International Journal of Computational and Experimental Science and ENgineering
(IJCESEN)
Vol. 11-No.3 (2025) pp. 6462-6472
http://www.ijcesen.com
ISSN: 2149-9144

Research Article

AI-Driven Data Automated Auditing and Governance Frameworks for Enterprise Data Engineering

Gayatri Tavva*

Rajeev Gandhi Memorial College of Engineering and Technology, Nandyala, Andhra Pradesh, India
* **Corresponding Author Email:** gayatr2i@gmail.com - **ORCID:** 0000-0002-5247-602X

Article Info:

DOI: 10.22399/ijcesen.3758
Received : 05 February 2025
Accepted : 28 March 2025

Keywords

AI-driven data governance
Automated data auditing
Enterprise data engineering
Machine learning in compliance
Data quality and lineage tracking

Abstract:

With the challenges of large-scale data evolving and enterprises battling with the complexities of the big-data ecosystem, there is growing urgency in the end-to-end intelligent and automated data auditing and governance systems. The conventional governance arrangements have been relatively inelastic, rigid, and unaware, which makes them inadequate to deal with changing and nonhomogeneous data spaces. This review is an insight into how artificial intelligence (AI) overhauls the prospect of next-gen data governance and auditing in enterprise data engineering. AI-enabled systems deliver ongoing quality, lineage, policy enforcement, and regulatory compliance by using machine learning, natural language processing, and knowledge graph technology to examine metadata and data lineages and the policies associated with them. These technologies improve decision making, cut down any human error, and provide an opportunity to predict anomalies in data and risks of accessing information. The article addresses the use of predictive auditing, automation of the interpretation of policies, and context-aware access control enabled by AI. It also lists some of the critical implementation issues as biased data, the explainability of the models, and the intricacy of involving the system. Moreover, the review details the role that federated and reinforcement learning play in the context of safe and flexible governance of distributed systems. With the introduction of AI and data governance together, the future of enterprise data engineering is bringing the missing pieces of transparency, compliance, and operational resilience. The paper summarizes the existing studies and provides a guideline for implementing AI-driven governance architectures to achieve sustainable, compliant, and efficient data management behaviors

1. Introduction

In a modern business where the data-first approach reigns supreme, enterprises have become more dependent on large, complex, and heterogeneous data infrastructures as a source of strategic decision-making, regulatory compliance, and operational intelligence at all levels of enterprise activity. This dependency is because contemporary businesses capture and produce massive volumes of data that have never been seen before through various systems, including IoT, cloud, customer interfaces, internal business systems, and third-party integrations. The volume, variety, and velocity of such an influx of data, known colloquially as the three Vs, has now become too great to fit the customary methods of data auditing and governance. Existing legacy systems that rely heavily on manual control, pre-determined sets of rules, and fixed work in processes cannot cope with the dynamism and intricacy of real-time data operations. These traditional methods are usually not flexible enough to support large data ecosystems and the dynamically changing regulatory environment, and lack the flexibility (distributed, multi-cloud, and hybrid data environments) [1][2].

To overcome such constraints, artificial intelligence (AI) has come in, a revolutionary technology with its advanced features that changes data audit and governance in the ecosystems of the enterprises. AI-enabled data auditing entails the implementation of smart algorithms that will run continually to observe, analyze, and certify the integrity and quality, compliance, lineage, and utilization of the enterprise data assets [3]. Such algorithms will detect inconsistencies in data, detect anomalies, validate the schema compliance, and ensure the integrity and transparency of data transformation pipelines during data ingestion to insight [4][5]. Parallel to these, data governance covers the larger scopes of setting up policies, imposing standards, and access control to secure and optimally utilize the usage of data throughout the organization. Once the AI is implemented on these governance systems, it enhances their power by allowing proactive surveillance, real-time changes to rules, and fully automatic enforcement of data regulations, as the level of manual control and rectification is greatly minimized [6][7].

Furthermore, the interaction of AI and enterprise data engineering will lead to the establishment of closed-loop data ecosystems possessing the capability of self-regulation. These systems automatically support data credibility, integrity, privacy, and compliance with regulations in an ever-changing digitized world. It is especially essential in industries with strict regulation, like healthcare, finance, defense, and telecommunication, where a lack of compliance may lead to serious legal, fiscal, and reputational consequences [2][4]. The capacity to audit and control data in real time is not a supplement in such industries but a regulatory and operational necessity.

Such an overview plunges into the present situation and emerging possibilities of AI-based information-auditing and governance models. It looks at the architectural foundations that support these systems, the fundamental technology that makes them work (such as machine learning (ML), natural language processing (NLP), and graph analytics), and what real-world difficulties operate surround organizations when realizing such systems. These obstacles come in the form of algorithmic bias, integration complexities, explainability, and governance in the decentralized environments [8][9][10]-[13]. In addition, the review also points out the impact of the international data-privacy laws and compliance requirements, which include the General Data Protection Regulation (GDPR) and California Consumer Privacy Act (CCPA), as the primary factors that accelerate the implementation of AI-based governance systems. It also analyses the trends and patterns in the usage of these technologies, and provides insight into how AI is changing the future of enterprise data engineering in terms of smart, automated, and scalable governance.

2. The Role of Artificial Intelligence in Modern Data Governance

Artificial intelligence (AI) has become a disruptive element in the development of data governance in enterprises, such that it is overhauling the modes through which corporations approach data management, all the way to how data is monitored and secured. The more traditional governance mechanisms, dependent on the use of fixed rules, manual control, are being complemented by AI-based mechanisms that have the ability to take in historical patterns of data, identify deviations, and automatically update governance mechanisms based on the new insights. The central premise underlying this change is the fact that AI, especially machine learning (ML) models, can change automatically to keep up with the shifting data landscape. One can condition these models to extract patterns in access, sharing, and transformation of data, so that sensitive levels of information can be labeled, violations of various policies highlighted, and remediation suggestions can be produced as automated conditional actions that satisfy regulation and organizational-specific policies [1, 15-17]. The transition decreases man-driven reliance and improves effectiveness and uniformity in the processes of governance.

Metadata analysis is one of the landscapes that make AI-driven governance possible. Metadata, or data about data, is the information that is essential to understand the origin, organization, and history of transformation of datasets. By means of their sophisticated analytical functions, AI systems can process metadata at scale, meaning that it is possible to accurately define data lineage and provenance. Such functionality is crucial in gaining end-to-end visibility in the data flows, and this, in turn, is the prerequisite to establishing compliance with more demanding data protection regulations such as the General Data Protection Regulation (GDPR) in the European Union and the California Consumer Privacy Act (CCPA) in the United States. AI allows not only tracing but also strengthens auditability, as its automated metadata surveillance produces logs of data usage and manipulation that are immutable, and thus have significant value in compliance reviews as well as investigations of forensic interest [2,15-20].

Another step taken to broaden the applicability of AI in government is natural language processing (NLP). Machines that can comprehend, interpret, and process human language are made possible in this sub-branch of AI. In this light, the NLP-enabled tools may read and recognize creepy clauses in law documents, in corporate policies, service-level agreements (SLAs), and regulatory texts, converting them to machine-readable governance rules. This feature also removes the historical dependence on legal and compliance staff to interpret complex records by hand, for a faster and safer compliance process, and eliminates the risk of making an error by people reading handwritten material [3]. As an example, an AI tool, based on NLP technology, can read one of the GDPR points and automatically adjust data access controls within the company to support its stipulations in different departments. Data access governance can also be heightened using AI-intelligent enforcement policies like role-based access control (RBAC) and dynamic data masking. The AI systems will enable an automatic recommendation or even real-time access adjustments based on regular user behavior

Table 1. *This table presents a condensed benchmark of key data quality and compliance risk dimensions, highlighting industry-standard performance targets, applicable data protection regulations (e.g., GDPR, HIPAA), and commonly used tools that support governance and regulatory compliance across sectors. Block Diagrams and Proposed Theoretical Model. The following block diagram presents the primary elements ofoptimizing user experience in AI-fueled enterprise sales tools, prioritizing seamless integration and user interaction.*

Dimension	Benchmark/Target	Key Regulation(s)	Suggested Tool(s)
Accuracy	≥ 95%	GDPR, HIPAA	Talend, Informatica
Completeness	≥ 98%	GDPR (Art. 5), HIPAA	BigID, Collibra
Consistency	100% uniform data	ISO 27001	AWS Glue, Informatica
Timeliness	Real-time or ≤ 1 day	HIPAA, GDPR	WhyLabs, Google DLP
Validity	100% format match	GDPR, ISO 27001	Talend, Collibra
Provenance	100% traceability	GDPR (Art. 30), SOC 2	BigID, IBM Watson OpenScale
Consent Records	100% user consent	GDPR (Art. 7), CCPA	OneTrust, TrustArc
Auditability	Logs within 24 hrs	HIPAA, PCI-DSS	Azure Monitor, Google Audit Logs

observation, access logs, and usage patterns to make sure that only authorized users touch sensitive data. Additionally, AI can detect deviations or other suspicious activity, e.g., unusually accessing restricted data sets or downloading data at unpredictable times, and react to them in real time, perhaps simply by masking the data, sending an alert, or executing an automated analysis. Such determinative services are especially vital in distributed enterprise scale settings, whereby it is unviable to manually manage activity on access to data in terms of departments, geographical location, and platform [2-7, 15-18].

Besides access control, the ability to detect anomalies is yet another intensive usage of AI when it comes to proactive data governance. Based on unsupervised learning methods and statistical models, AI can detect anomalies and inconsistencies at all phases of the data lifecycle. They could involve inconsistencies across data transformation operations, schema alteration surprises, or increases in data access loads. In contrast with traditional systems based on pre-determined rules and thresholds, AI models learn to dynamically change baselines and thus would better respond to subtle and context-specific anomalies [4]. To elaborate, assume that a particular dataset, which is frequently accessed by a finance team or department, is accessed one day by a third-party contractor by a query, the AI model can raise this as a potential breach, thus limiting unauthorized access or leakage of information before such can escalate into a full-blown security breach [6-10, 16-19].

The total effect of these AI functionalities gives birth to the paradigm shift in governance, which is reactive to proactive. Instead of the violations being identified during the periodic audit or manual inspection, AI systems can be used to enforce the policies in real-time and act quickly in response to threats. The transformation not only increases data security and compliance but also enables organisations to deal with their data more effectively and reliably. With greater regulatory pressure and ever-growing quantities of data, AI-powered governance is proving to be an essential component of contemporary enterprise data strategies, a scalable, smart, and flexible approach to preserving the most valuable digital asset an organization possesses [14-18].

Table 2. *Key roles and benefits of artificial intelligence in enhancing modern data governance practices.*

AI Capability	Function in Data Governance	Benefits
Data Classification	Automatically categorizes data based on content and context	Enhances discoverability and access control
Anomaly Detection	Identifies irregularities or potential data breaches	Improves data integrity and security
Metadata Enrichment	Auto-generates tags, descriptions, and context from raw data	Increases metadata accuracy and usability

AI Capability	Function in Data Governance	Benefits
Policy Enforcement	Implements data handling rules using AI-based decision logic	Ensures regulatory and organizational compliance
Predictive Auditing	Anticipates potential compliance violations based on historical patterns	Enables proactive governance and risk mitigation
NLP for Regulatory Parsing	Translates legal and compliance documents into actionable rules	Reduces manual interpretation effort
Lineage Tracking	Maps data flow and transformation across systems	Supports transparency and accountability
Access Monitoring	Uses AI to detect unusual access behavior or privilege misuse	Strengthens access control and auditing
Data Quality Monitoring	Continuously scans for missing values, outliers, or format mismatches	Maintains high data quality standards
Consent Management	Automates consent tracking and enforcement for data usage	Enhances compliance with privacy regulations (e.g., GDPR)
Risk Scoring	Quantifies risk based on data usage, sensitivity, and user behavior	Prioritizes governance focus based on risk levels
AI-Driven Recommendations	Suggests governance policies or remediation actions based on trends	Accelerates policy development and adaptation
Knowledge Graphs	Links data, users, and policies semantically to provide contextual insights	Improves decision-making and traceability
Explainable AI (XAI)	Provides rationale behind AI decisions in governance models	Enhances trust and auditability in governance actions

3. Intelligent Data Auditing: Automating Trust and Traceability

In recent years, the notion of data auditing has changed dramatically as it has evolved towards constantly moving, real-time tracking and predictive diagnostics that are supported by artificial intelligence (AI). The transformation has been supported by increased complexities, volumes, and sensitivity of enterprise data, which has to be consistently monitored to maintain accuracy, compliance, and credibility. AI-based auditing systems have already reached the constant, routine mode, constantly scanning the data streams and pipelines, searching for anomalies, inconsistencies, and duplicates, and identifying possible issues with the governance rules. This takes the form of a proactive approach in that problems are caught in real-time and rectified immediately, so there is a marked decrease in the time by which the problem is spotted and fixed. Graph-based AI models used to audit the interrelated and sophisticated data environments can be deemed one of the most innovative developments in this field. Graph-based models also enable the representation and analysis of complex interconnections among varied data objects, users, systems, and processes, unlike the traditional linear or tabular methods. These models allow the real-time tracking of dynamic data lineage, which lets enterprises visualize and trace the revision of data through multiple applications, as well as databases and transformations [5, 7, 13, 16, 19, 20]. Such visibility is imperative in keeping data ecosystems transparent and integrity intact, but also when it comes to conducting root cause analysis on anomalies or violations that may arise. Knowing the way in which a data element has been created, changed, and used throughout its lifecycle, organizations can be in a better position to regulate governance policies, enable compliance reporting, and reduce the risk as well.

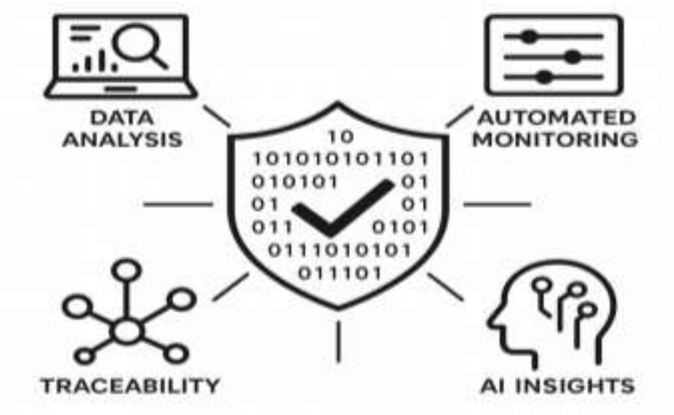
The other major development is the development of predictive auditing, which is known as the forward-looking, intelligence-based risk management approach. By use of supervised machine learning algorithms, AI can be trained using historical audit logs and operational traces to learn about patterns associated with prior incidences of non-compliance or data breaches, or operational interference. Such trained models will then be capable of detecting the possible areas of concern before they develop into systemic failures [6, 8, 12, 14, 18]. As an example, say a particular kind of access or data manipulation is known to historically occur along with policy violations, the predictive AI can imminently indicate such an occurrence on the fly and issue alerts, or cause automated remediation to occur. This not only increases operating resilience but also transforms the paradigm of governance from reaction to prevention, which gives an organization a large strategic advantage.

One of the crucial ancillary components of contemporary AI-assisted auditing is the feature of intelligent data profiling and the automation of the analysis of datasets, which helps to evaluate their structure, consistency, and quality. In order to assess the compliance with the schema, determine missing values (nulls), outliers, and find formatting errors or unusual data patterns, AI tools can quickly scan the datasets [7, 12, 14, 20]. Data health of these automated profiles gives organizations visibility into the health of their data on an ongoing basis in order to create a real-time data quality dashboard and alerts to keep the data stewards updated on arising concerns. This is a real-time feedback loop that enables better decision making and helps in perpetual improvement efforts through data management practice.

The capability of current AI technology to generate explainable audit trails is also significant, as this type of expert guidance is essential to proper transparency and accountability in the process of automated decision making. As opposed to traditional audit records that merely store access history and user activity, AI-generated audit logs that provide explanations for decisions about data governance usually consist of explainable model predictions [13, 14, 17,19]. As an example, an explanation may be given along with closing an AI model when it blocks access to a sensitive set of data or marks a transaction as suspicious, i.e., as not within the bounds of known user behavior or policy-derived limits that can be used to refute it. This explainability is not only essential to retain the confidence of the stakeholders but also helps legal defensibility during audit, regulatory investigations, and external review. The very possibility of explaining and justifying AI-driven governance actions becomes a non-negotiable requirement as more legislations, such as GDPR and CCPA, continue to drive transparency and accountability across automations.

Conclusively, data auditing has grown tremendously with the incorporation of powerful AI technologies by overcoming the traditional data audit that dates back in time and queries on more complex, real-time auditing as predictive and explainable audits of enterprise data ecosystems. Whether it is via graph-based lineage tracking and predictive modeling, to the remarkably smart data characterization and interpretable audit trail, AI is re-making the very nature of how companies track, control, manage, and protect their data assets. This is a revolution that enables businesses to be in a position to sustain greater standards of compliance, enhance their efficiency in operations, and foster more confidence in their data governance regimes.

Figure 1. *Intelligent data auditing framework*



highlighting automated monitoring, AI insights, data analysis, and traceability for enhanced trust.

4. Enabling Technologies: Machine Learning, NLP, and Knowledge Graphs

A combination of several advanced technologies to improve intelligence, flexibility, and scalability in an enterprise data environment is the foundation that makes AI-driven auditing and a governance framework effective and viable in the long term. The most important of these technologies includes machine learning (ML), which is the analytical backbone of a significant portion of automated auditing and governance systems. Supervised and unsupervised ML techniques have been used to create models evaluating data quality as well as classifying datasets depending on their sensitivity, identifying anomalies, and making predictions of possible instances of non-compliance. Supervised learning is based on labeled past data that is used to predict what can happen in the future like access violations or degraded data quality and unsupervised learning can be used to discover hidden patterns and clusters within unstructured data, and yield insight on compliance risks that may be otherwise unseen or on inconsistencies in how data is used [1]. Such capabilities ensure that ML would be instrumental in helping organizations uphold a high level of data integrity, security, and regulatory compliance on a vast network of data pipelines.

Also fundamental to AI-enabled governance is Natural language processing (NLP), which can be used to solve one of the major bottlenecks in conventional governance procedures, the manual analysis of textual policies, regulatory documents, legal contracts, and service agreements. NLP algorithms are able to read and semantically parse this human-readable text, extract applicable statements of clauses and constraints, and translate them into machine-readable rules that may be systematically applied to data processes. This automated translation makes sure that the smooth governance policies are in place, they are exposed, and there are ways to enforce them as regulatory standards increase. To give an example, an NLP may analyze a GDPR compliance document to find statements about consent and retention of user data and map them to specific controls in a data management system automatically [8]. This will save human effort dramatically, eradicate error in interpretation, and accelerate compliance procedures.

Knowledge graph, an example of a foundational technology, is the knowledge graph, a structure designed to tie data elements, systems, and users into an interconnected graph-structured semantic relationship and ontologies. These graphs allow a complete, more contextualized overview of the data ecosystem in which it is simpler to query connections, guess absent connections, and comprehend the heritage and lifecycle of data resources. An example would be that a knowledge graph would be able to trace the path of personally identifiable information (PII) as it flows through source systems, through to analytical models, showing all transformations and places of access in between. This facilitates the granularity tracking of lineage and policies, which are fundamental in regulation reporting and internal auditing [9]. Through the inclusion of ontological logical thinking, knowledge graphs can also perform inferencing, the ability of the system to learn new facts, and even apply obscure governance rules not necessarily programmed in.

Federated learning has proven as a good solution to improving data governance where security and privacy of data are what matter most. Federated learning does not limit the centralized approach to training like the centralized methods do, as it enables decentralized data providers to construct ML models together without the need to exchange the raw data. The method is especially beneficial in business sectors like healthcare and finance, where industry regulations limit the ability to transfer sensitive information across geographical or institutional borders. Local training and aggregation of model updates instead of data in federated learning enables privacy-preserving, collaborative model development. It is GDPR and HIPAA-compliant, so it can be useful in a privacy-respectful management of large data masses [10].

Moreover, reinforcement learning (RL) is gradually becoming discussed as a flexible and adaptive policy optimisation process in data governance regimes. In comparison with supervised learning, where fixed data is used, RL algorithms learn as a result of interaction with the environment (taking actions, observing consequences, refining policies depending on feedback). RL can also be used in the domain of data governance, where it can be used to continually optimize access control rules, policy thresholds, and anomaly detection models with respect to continually updated feedback based on system behavior and user interactions. As another example, in a case where one particular policy generates false alerts all the time, the RL model can readjust its parameters in a way that reduces the amount of overhead without sacrificing security. This kind of self-optimizing government is proven to be especially valuable in large-scale, fast-moving data environments where fixed regulations can soon be outdated [11].

These technologies together constitute a synergistic platform that allows AI-enabled auditing and governance systems to graduate out of the boxy, rule-driven platform into intelligent, self-directed platforms that have the ability to run without any supervision. Machine learning guarantees constant pattern recognition and predictive behavior; NLP unlocks an interpretation and execution of increasingly complicated human-composed policies; knowledge graphs facilitate greater contextual awareness and traceability; federated learning enables privacy-sufficient multi-party collaboration; and reinforcement learning brings ongoing adaptation and policy adaptation. Combining these capabilities, the enterprises are in a better position to handle data within the distributed environment, provide real-time governance compliance, and drive trust in the data systems, bringing the technical governance capacities in line with the business strategic objectives.

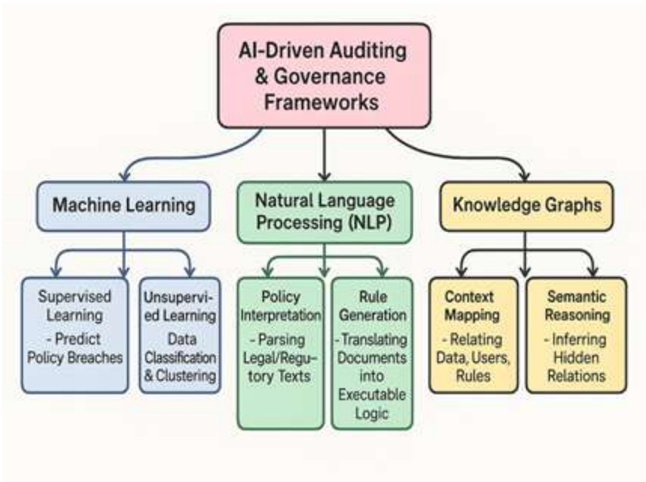


Figure 2. *AI-driven auditing and governance framework integrating machine learning, NLP, and knowledge graphs for automated compliance and decision-making.*

5. Implementation Challenges and Risk Considerations

As much as the AI-powered systems of audits and governance have a promising potential to transform the enterprises' management of data in general, the possibilities for the active implementation of these systems in real-world situations still have a few critical points that remain a limiting factor impacting the policies and efficiency of the practices themselves. Among those of the most burning ones is the issue of information bias that occurs when the models of machine learning are trained on the datasets that are incomplete, unbalanced, or they are not representative of the actual operating environment. This sort of bias can have devastating effects, such as sensitive data being misclassified, improper access privileges granted, as well as poor choices of policy enforcement. In industries that are controlled, these errors can cause not only operational inefficiencies but also legal penalties. The origin of this issue lies in the fact that AI models, especially supervised learning models, require the quality and diversity of training data. In case historical data is structured with hidden human biases or systematic inconsistencies, then these may be duplicated and, in the worst case, larger than in human decision-making activity [12].

The second significant challenge is the problem of explainability of models. Most AI models, especially the ones that utilize deep learning architectures, can display excellent accuracy but tend to be rather opaque, acting as a black box and producing outputs in a form that cannot be openly regarded as opaque and indecipherable instructions. This lack of transparency is particularly objectionable in areas of high magnitude like healthcare, banking, and legal compliance, when the decisions should be auditable and defensible. Regulatory bodies are growingly requiring automated decision-making tools to offer human-interpretable explanations of their results. In these scenarios, a decision model designed by AI is unable to explain its actions, which not only hinders the trust of the stakeholders but can also go against the regulatory provisions of the General Data Protection Regulation (GDPR), which considers automated profiling and right to explanation [13].

Another overwhelming hurdle is that of integration complexity. The majority of large businesses work in an unbalanced data landscape consisting of legacy frameworks, public clouds, outside instruments, and hundreds of forms and unorganized information sources. Sprinkling AI-based governance solutions in these various platforms will involve a massive architecture redesign. Interoperability has to be guaranteed on many layers, not only with compatibility of the data schema, but also the communication protocols between systems, and the API stack, and may necessitate the involvement of cross-functional groups, a considerable monetary investment, and a potentially long-term implementation. Moreover, AI models will have to be seriously customized to suit each enterprise data management workflows, business logic, and the domain-specific governance standards; a one-size-fits-all solution is not possible [14-15].

Along with technical obstacles, regulatory convergence is not a fixed target. The regulation regarding privacy and use, and retention of data is ongoing, and international differences in regulation compound the complexity. As an example, GDPR in Europe, CCPA in California, and future regulations in other regions all have varying requirements of compliance. AI-based systems within any governance system should therefore be flexible enough to support even these changes. This brings the use of continuous model lifecycle management frameworks with retraining, validation, version control, and compliance test capabilities. In the event that these controls are not established, it is risky that organizations will implement outdated or non-compliant AI logic, which might have dire effects on them in terms of legal repercussions and reputational issues [15-17].

Finally, human monitoring is an absolute necessity even in the most artificial intelligence-supported space. Although AI is especially well-matched to automating repetitive tasks, detection of patterns, and expansion of governance activities, it lacks the subtle decision-making and detailed background that human data stewards have. It is often wise to seek options on the sensitivity of data, ethical use of data, or cross-department sharing of data; at times, these decisions need a rich understanding of organizational context, stakeholders, or the long-term impact of the business areas where AI still remains inefficient. Therefore, the AI role needs to be regarded as augmentative as opposed to substitutive. The most effective implementation patterns unite the operational capacity of AI with a deeply entrenched knowledge base and ethical consciousness of human factors, resulting in a blended, governance-like model that can maintain a balance between automation and responsibility [18-20].

Table 3. *This table showcases real-world implementations of AI governance applications across critical industries.*

Sector	Case Study Description	AI Tool/Technology Used	Outcomes/Impact
Healthcare	Deployment of AI models for breast cancer detection through mammography screening. Deep learning algorithms were trained on diverse datasets across the UK	Deep Learning (Convolutional Neural Networks - CNNs)	Enhanced diagnostic accuracy, early detection of cancer, reduction in false positives and

Sector	Case Study Description	AI Tool/Technology Used	Outcomes/Impact
	and the USA and matched or outperformed expert radiologists.		negatives, and less radiologist workload.
Retail	Adoption of AI for personalized customer experiences through recommendation systems in online platforms. This supports product targeting and dynamic pricing strategies.	Machine Learning-based recommender systems and customer behavior prediction models	Improved user engagement, increase in sales conversion rates, higher customer retention, and optimized marketing campaigns.
Finance	Financial institutions using AI for fraud detection and credit scoring. AI algorithms analyze transactional patterns to flag anomalies and provide explainable credit risk models.	Anomaly detection models, Explainable AI (XAI), supervised and unsupervised learning algorithms	Reduced fraud rates, faster transaction approval, better credit risk transparency and compliance with financial regulations.

6. Future Outlook and Enterprise Impact

The maturity of AI-enforced auditing and governance is currently quickly evolving to the generation of autonomous sets of data engineering whose capacity for self-monitoring, self-diagramming, and self-correction is minimal. These smart systems are made in such a way that they constantly monitor the status of data assets, identify inconsistencies and eliminate errors on the fly, and make sure that data management fits the organizational policy and regulatory frameworks. With an expanding number of enterprises going to the utilization of modern architectural techniques like data mesh and data fabric, which distribute data ownership and advance domain-specific data stewardship, the problem of relatively steady and authoritative governance increases in its scope exponentially. In its turn, the governance structures should be re-designed to be scaled out along the border of distributed teams, technologies, and environments. In this respect, artificial intelligence plays the most important role as a solution to connect abstract governance policies on the one hand and their implementation into the mixed-data environment on the other. AI systems not only make data accurate, secure, accessible, and compliant throughout the data life, but also at all times and locations of its processing.

Moreover, the most advanced technology providers are integrating AI-native management capabilities into the data infrastructure of the most critical frameworks, such as data lake houses, data catalogs, and master data management systems. These inherent features enable metadata to be managed automatically, identify anomalies, classify sensitive data, and apply data access policies based on real-time use. The automation of governance is bound to be automated further with the incorporation of generative AI technologies. Through the use of transformer-based architectures and large language models, companies can generate governance documentation, compliance reports, audit logs, and even regulatory plans responsive to the distinct data schema and business needs of companies by training a program on both development and production data. These systems are also able to read the natural language regulations and turn them into working governance rules, mitigating the reliance of the legal and compliance teams to interpret the regulations manually.

Finally, there is an assemblage of AI, data governance, and enterprise data engineering, which are redefining the future of organizational data management. The result of this fusion will be intelligent, responsive, and scalable ecosystems in which compliance, accountability, and the quality of the data is managed proactively and no longer the result of a reactive effort. In this type of ecosystem, data is not only recognized as an asset, but it is managed as a critical data resource with legal, ethical, and operational ramifications. With technologies still in their infancy still having the potential to transform the standards of enterprise responsibility, organizations will be able to create data infrastructures that are effective, auditable, resilient, and aligned with business aspirations as well as societal expectations.

Table 4. *This table compares major AI governance platforms across key criteria, including features, strengths, limitations, and optimal use cases, helping users evaluate appropriate tools based on infrastructure, compliance needs, and governance maturity*.

Platform/Tool	Developer/Provider	Key Features	Strengths	Limitations	Best Suited For
IBM Watson OpenScale	IBM	- Bias detection- Explainability- Drift monitoring-	- Integrated with IBM Cloud- Supports model	- Primarily optimized for IBM ecosystem-	Enterprises using IBM Cloud and

Platform/Tool	Developer/Provider	Key Features	Strengths	Limitations	Best Suited For
		Model accuracy tracking	lifecycle monitoring- Strong explainability features	Complex setup for non-IBM environments	seeking end-to-end AI lifecycle governance
Microsoft Responsible AI Dashboard	Microsoft	- Fairness, error, interpretability, and counterfactual analysis- Model monitoring- Integration with Azure ML	- Open-source components- Integrates with Azure and Azure ML pipelines- Interactive dashboards	- May require customization for non-Azure models- Limited support outside Python-based models	Organizations developing AI within Azure ecosystem
Google Vertex AI Model Monitoring	Google Cloud	- Continuous model monitoring- Drift and skew detection- Integration with Vertex AI pipelines	- Strong integration with Google Cloud- Real-time monitoring- Scalable architecture	- Limited explainability features compared to others- Best for Google ecosystem only	Businesses operating within Google Cloud infrastructure
AWS SageMaker Clarify	Amazon Web Services	- Bias detection- Explainability with SHAP values- Data drift monitoring	- Strong integration with SageMaker pipelines- Scalable and cloud-native- Comprehensive metrics for fairness	- Lacks visual dashboards compared to peers- AWS-specific; less flexible for hybrid environments	AI solutions developed and deployed on AWS
AI Fairness 360	IBM (Open Source)	- Bias detection for data and models- Over 70 fairness metrics and mitigation algorithms	- Open-source and customizable- Rich fairness toolkit- Compatible with Jupyter notebooks	- No GUI (code-heavy usage)- Requires technical expertise	Academic researchers and data scientists needing fairness audits
Fiddler AI	Fiddler Labs	- Explainability- Bias and drift monitoring- Customizable alerts- Real-time dashboards	- Model-agnostic- Strong visualizations- Supports multiple deployment environments	- Commercial product with subscription costs- May need integration effort	Enterprises needing visibility into model performance and decisions
Truera	Truera	- Model explainability- Fairness and performance analysis- Pre-deployment and post-deployment testing	- Strong diagnostic tools- Integrates with CI/CD pipelines- Works with common frameworks (e.g., Scikit-learn, XGBoost)	- Subscription-based- Focused more on explainability than full lifecycle	Model developers and compliance teams focused on explainability and fairness
WhyLabs AI Observatory	WhyLabs	- Model monitoring- Data quality assurance- Drift and anomaly detection	- Cloud-native with integrations for ML pipelines- Automated alerts and reporting- Open-source support via WhyLogs	- Less emphasis on fairness and explainability- Requires some setup overhead	MLOps teams focused on production monitoring and data quality
Hazy	Hazy	- Synthetic data generation- Privacy-preserving AI governance- Data simulation	- Focus on data privacy- Strong GDPR compliance support- Enables model testing on synthetic data	- Niche use-case (synthetic data)- Not a full AI governance suite	Organizations requiring secure AI training with synthetic, privacy-compliant data

Platform/Tool	Developer/Provider	Key Features	Strengths	Limitations	Best Suited For
		and compliance tools			
Credo AI	Credo AI	- AI governance framework- Risk assessments- Compliance tracking- Policy management	- Policy-based approach- Good for audits and compliance- Stakeholder-centric dashboards	- Newer in the market- Less technical than other tools	Compliance officers, risk management teams, and policy leaders

Conclusion

The use of AI in data auditing and data governance programs can be discussed as a paradigmatic shift in the way organisations control the information, monitor, and secure it. Such systems present an effective alternative to the manual, fixed governing methods, providing flexible, expandable, and intelligent control over complicated data ecosystems. AI can boost anomaly detection, apply policies, data lineage, and regulatory compliance in real-time due to machine learning, natural language processing, and graph-based modeling. These capabilities are becoming essential in a period of fast-paced data development, tighter privacy policies, and a high cost of data leakage and non-conformance.

Nonetheless, there are issues surrounding the use of AI-based frameworks, including bias associated with algorithms, issues of integration, and explainability. A way to deal with them is through intensive model governance, alignment of stakeholders, and ongoing observation of AI systems to achieve fairness and accountability. As companies migrate to data mesh and data fabric architecture, the pressure on decentralized, autonomous data governance solutions will rise.

Finally, AI-driven data governance and auditing turn the data engineering side of the organisation into a more reliable, compliant, and future-proof operation. Through the adoption of such frameworks, organizations will be able to maximize both data trust and data transparency, but they will get some strategic benefits as well, boosting decision-making and the efficiency of their operations.

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.

- **Author contributions:** The authors declare that they have equal right on this paper.

- **Funding information:** The authors declare that there is no funding to be acknowledged.

- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

[1] Wamba-Taguimdje, S. L., Wamba, S. F., Kamdjoug, J. R. K., & Wanko, C. E. T. (2020). Influence Of Artificial Intelligence (Ai) On Firm Performance: The Business Value of Ai-Based Transformation Projects. *Business Process Management Journal*, 26(7), 1893-1924.

[2] Zarsky, T. Z. (2016). Incompatible: The Gdpr In the Age of Big Data. *Seton Hall L. Rev*., 47, 995.

[3] Shollo, A., & Galliers, R. D. (2016). Towards An Understanding of the Role of Business Intelligence Systems in Organisational Knowing. *Information Systems Journal*, 26(4), 339-367.

[4] Kandolo, W. (2024). Ensuring Ai Data Access Control in Rdbms: A Comprehensive Review. In Proceedings of the Ieee/Cvf Conference On Computer Vision and Pattern Recognition (8400-8407).

[5] Malik, N., & Malik, S. K. (2020). Using Iot and Semantic Web Technologies for Healthcare and Medical Sector. *Ontology-Based Information Retrieval for Healthcare Systems*, 91-115.

[6] Haug, A., Zachariassen, F., & Van Liempd, D. (2011). The Costs of Poor Data Quality. *Journal Of Industrial Engineering and Management (Jiem)*, 4(2), 168-193.

[7] Otto, B. (2011). A Morphology of the Organisation of Data Governance.

[8] Dong, X. L., & Srivastava, D. (2015). Big Data Integration. *Morgan & Claypool Publishers*.

[9] Ehrlinger, L., & Wöß, W. (2016). Towards A Definition of Knowledge Graphs. *Semantics (Posters, Demos, Success)*, 48(1-4), 2.

[10] Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). Federated Machine Learning: Concept And Applications. *Acm Transactions on Intelligent Systems And Technology (Tist)*, 10(2), 1-19.

[11] Panigrahi, S., Nanda, A., & Swarnkar, T. (2020). A Survey on Transfer Learning. In Intelligent and Cloud Computing: Proceedings of Icicc 2019, Vol(1)(781-789). Singapore: *Springer Singapore*.

[12] Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A Survey on Bias and Fairness in Machine Learning. *Acm Computing Surveys (Csur)*, 54(6), 1-35.

[13] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). Why Should I Trust You? Explaining The Predictions of Any Classifier. In Proceedings of The 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining (1135-1144).

[14] Lakarasu, P. (2022). Ai-Driven Data Engineering: Automating Data Quality, Lineage, And Transformation in Cloud-Scale Platforms. *Lineage, And Transformation in Cloud-Scale Platforms* (December 10, 2022).

[15] Gudepu, B. K., & Eichler, R. (2024). The Role of Ai in Enhancing Data Governance Strategies. *International Journal of Acta Informatica*, 3(1), 169-187.

[16] Prasad, N., & Paripati, L. K. (2020). Ai-Driven Data Governance Framework for Cloud-Based Data Analytics. *Webology* (Issn: 1735-188x), 17(2).

[17] Onoja, J. P., Hamza, O., Collins, A., Chibunna, U. B., Eweja, A., & Daraojimba, A. I. (2021). Digital Transformation and Data Governance: Strategies for Regulatory Compliance and Secure Ai-Driven Business Operations. *J. Front. Multidiscip. Res*, 2(1), 43-55.

[18] Galla, E. P., Kuraku, C., Gollangi, H. K., Sunkara, J. R., & Madhavaram, C. R. Ai-Driven Data Engineering Transforming Big Data Into Actionable Insight. *Jec Publication*.

[19] Paleti, S. (2024). Data Engineering For Ai-Powered Compliance: A New Paradigm In Banking Risk Management. *European Advanced Journal for Science & Engineering (Eajse)-P-Issn 3050-9696 En E-Issn* 3050-970x, 2(1).

[20] Potdar, A. (2024). Ai-Based Big Data Governance Frameworks for Secure and Compliant Data Processing. International Journal of Artificial Intelligence, Data Science, And Machine Learning, 5(4), 72-80.