LTIMindtree

Point of View

Best Practices for Data Governance in the Age of AI

LTIMindtree

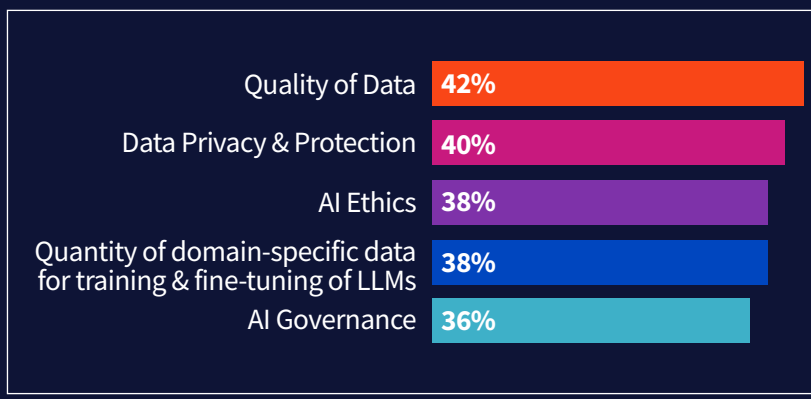
Bill Gates once said, ***"Soon after the first automobiles were on the road, there was the first car crash. But we didn't ban cars—we adopted speed limits, safety standards, licensing requirements, drunk-driving laws, and other rules of the road."***[1]

This analogy can be applied to the role of data governance in AI. Just like cars needed safety rules and driving standards when they were invented, the rise of AI demands the establishment of robust data governance practices. Data governance practices are like the "rules of the road", to ensure AI models are developed and deployed responsibly, ethically, and safely. By implementing robust data governance frameworks, organizations establish accountability, ensure data quality, secure data, and mitigate the risks associated with AI while harnessing its transformative potential. This is Data Governance for AI.

On the other hand, AI for Data Governance can significantly enhance data governance practices. It involves using Machine Learning (ML), Natural Language Processing (NLP), predictive analytics, and robotic process automation to streamline and enhance data governance practices. Before diving further, let us look at the key challenges related to data and AI governance.

Key challenges organizations face in the age of AI

Top Generative AI Challenges



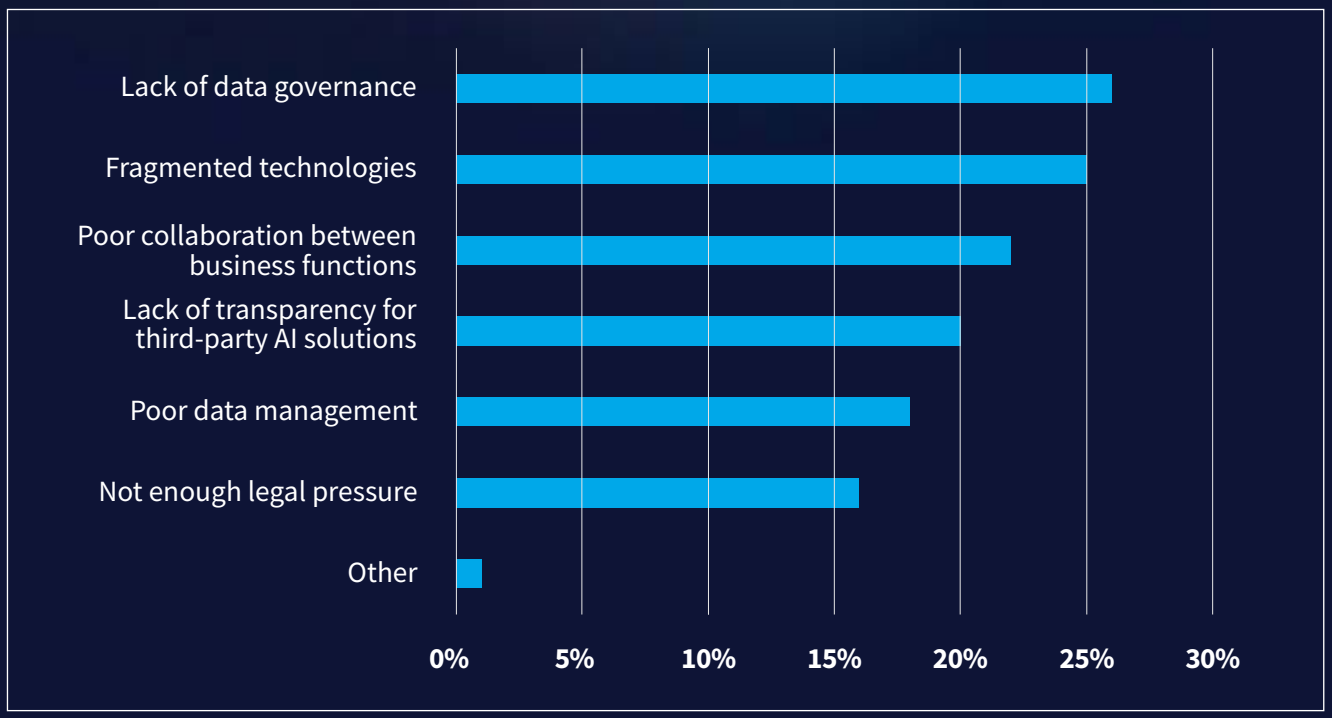
Informatica's CDO Insights 2024 report found that 36% of data leaders face challenges in AI governance.[2]

Source: Informatica's CDO Insights 2024 report

Figure 1: Top Generative AI challenges

A Gartner 2023 survey – AI Governance Frameworks for Responsible AI3 identified the biggest challenges to implementing AI governance as:

Challenges to Implementing AI Governance



Source: Gartner Report AI Governance Frameworks for Responsible AI (2023)[3]

Figure 2: Challenges to implementing AI governance

In the same report, some leaders highlighted the potential risks associated with the absence of Data & AI governance:

"Market will move exponentially; being late to the game will have a substantial impact."

- C suite, healthcare industry, 1-5k employees

"Poor business decisions are made as a result of lack of appropriate oversight."

- Director, manufacturing industry, 1-5k employees

"The greatest risk is that AI will be used improperly and result in improper or unethical outcomes."

- Director, educational services industry, 1-5k employees

"A system that shows bias cannot provide meaningful insights, and essentially becomes more nuisance than an effective tool."

- VP, manufacturing industry, 1-5k employees

Source: Gartner Report AI Governance Frameworks for Responsible AI (2023)[3]

Figure 3: Potential risk highlighted by leaders

We can summarize that the key challenges that organizations face in establishing governance include difficulty in ensuring data consistency, avoiding unreliable outcomes, and detecting and mitigating biases to ensure fairness. Additional challenges include navigating complex regulatory compliance, maintaining AI model transparency of AI models, establishing ownership and accountability, and prioritizing investment in technology and training.

With the rapid advancements in AI, it is important to establish a complete set of guidelines and best practices for data and AI governance. Let us explore these guidelines and best practices.

Key challenges organizations face in the age of AI

According to Amit Walia, CEO of Informatica, " Gen AI is here to raise us up, not replace us altogether. Regulations and guardrails will be a foundational part of that conversation, and as business leaders, we should not shy away from them. Instead, acknowledge their approach, participate in the dialogue, educate our teams, and embrace the revolution."

This balanced view highlights the proactive and collaborative approach to data governance, which ensures that AI's benefits are harnessed responsibly and sustainably.

Data acts as the fuel for AI models, enabling models to learn, self-correct, and provide more accurate insights. Therefore, it is crucial to establish Data and AI governance to ensure the reliability, consistency, and protection of the data.

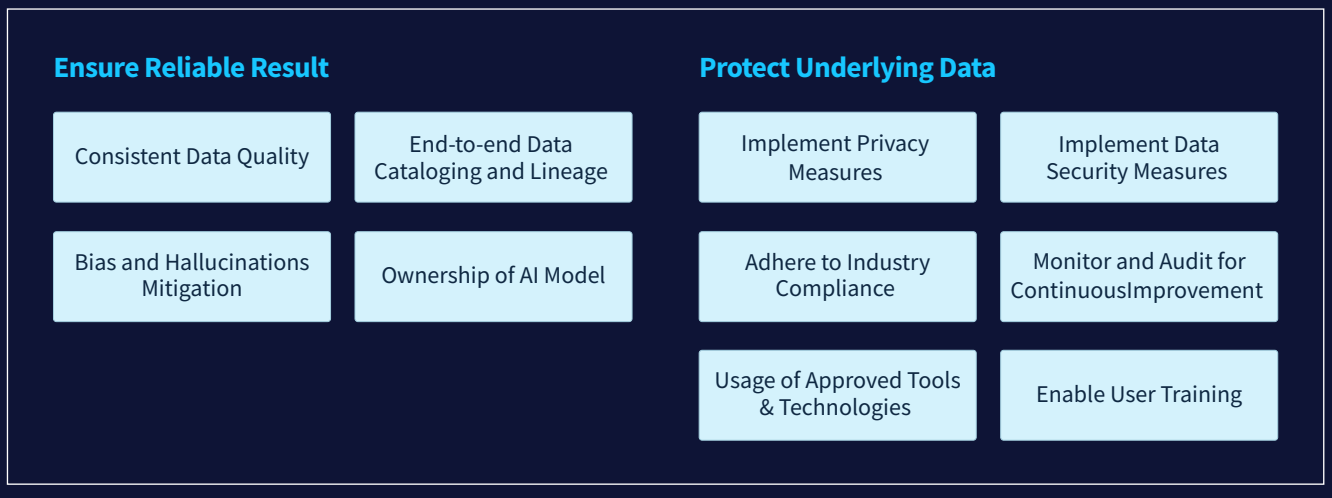


Figure 4: Fortifying Data Governance for AI

1. Ensuring reliable results from an AI model

Organizations often customize pre-trained models and perform prompt engineering, which requires access to internal data and external knowledge. Strong data governance prevents inaccurate or biased outcomes while maintaining reliability. For reliable outcomes from the AI model, ensure:

Consistent Data Quality: AI models are only as good as the data they are trained on, and hence, high-quality data is the foundation of reliable AI insights. For this, implement KPI-driven data validation and monitoring processes to maintain data accuracy, completeness, and consistency.

End-to-end data cataloging and lineage: Provide a clear and comprehensive view of data and pipeline lineage to help data scientists/users understand the meaning and context of the data they use to build an AI model. Also, maintain a catalog of AI models and versions centrally. This way, organizations can embrace transparency, which builds trust among stakeholders and users.

Bias and hallucination mitigation: Ethical considerations are inherent to data governance for AI. The governance framework must be designed to detect and mitigate biases in AI models to ensure impartiality and outcome consistency.

Ownership of the AI model: Embracing accountability is essential for managing the AI model and related data. Establish model ownership and stewardship (human involvement) to ensure AI-augmented solutions are reviewable, overridable, explainable, transparent, and compliant.

2. Protecting the underlying data

Another critical aspect of data governance is the protection of underlying data. AI interventions often involve accessing and processing sensitive/proprietary information and exposing confidential data and KPIs outside. To protect the data, ensure the following:

Privacy measures: Apply anonymization or pseudonymization to protect personal data. Implement privacy controls when extracting metadata and data from confidential files.

Data security measures: With the increased use of AI, protecting data from breaches and unauthorized access has become crucial. Implement strict access controls and encryption techniques at rest and in transit to ensure authorized access and data usage.

Industry compliance: Keep up with changing data protection and privacy regulations. Ensure adherence to relevant regulations and standards, such as General Data Protection Regulation (GDPR), Health Insurance Portability and Accountability Act (HIPAA), or industry-specific guidelines.

Continuous improvement: Regularly monitor AI models for performance and compliance with governance policies, conducting audits to identify and mitigate risks.

Usage of approved tools and technologies: Only tools that are legally approved by your organization should be used to develop AI models. All tools should be centrally cataloged.

Enable user training: Educate users on AI (especially generative AI or Gen AI) capabilities and limitations and establish ethical AI guidelines.

Best practices and guidelines: AI for data governance

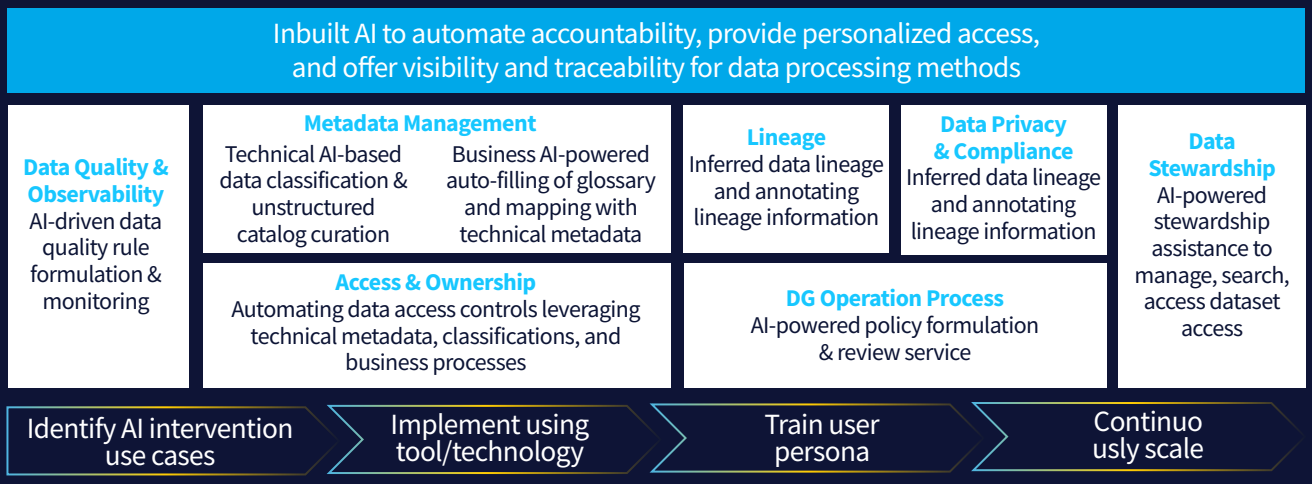


Figure 5: AI Strategies for Implementing Data Governance Core Processes

Invest in AI-powered tools and technologies: Develop a strategy to choose the right tool stack that clearly integrates AI and aligns with your organization's overall business goals and regulatory requirements. Invest in scalable, secure, and AI-powered technologies that not only handle varied forms of data but also become digital assistants working alongside the stewards.

Leverage AI-led data quality: AI intervention in data quality methodology can predict and detect anomalies before they can impact business outcomes. By utilizing AI-powered tools to identify patterns and issues in data, stewards can take proactive action to review and resolve these issues.

Enhance data discovery and classification: One of the significant challenges in data governance is the identification and classification of vast amounts of data across the enterprise. AI-powered data catalogs use AI/ML algorithms to catalog unstructured data, provide suggestions for auto-filling business glossaries, and provide classification and metadata linking recommendations. This automated process improves data discovery and classification and ensures comprehensive visibility into the organization's data.

Use AI to autogenerate data lineage: Understanding data lineage and its impact on downstream systems and processes is crucial for effective data governance. AI intervention can identify unknown links or hidden sources based on patterns, relationships, and dependencies. It can also append them to the data lineage for an end-to-end view of data transition across the value chain.

Embed AI in data privacy and compliance: Streamline privacy and compliance processes using automated data masking, risk scoring, and compliance monitoring.

AI-powered data access management: Automate data access controls by leveraging technical metadata, classifications, and business processes. This approach makes it easier for companies to secure data and automate access as they expand.

Empower data stewards with AI: Use AI assistance to manage, search, access, and converse with data assets, providing them with the necessary authority and resources to oversee them effectively. This, in turn, improves their day-to-day operations.

Apply AI for intelligent review and continuous improvement: Leverage AI to review industry-specific norms, implement data policies and privacy controls, and monitor KPIs to provide recommendations for improving the overall data governance process and standards.

How Intelligent Data Governance & Catalog (IDMC) drives AI intervention

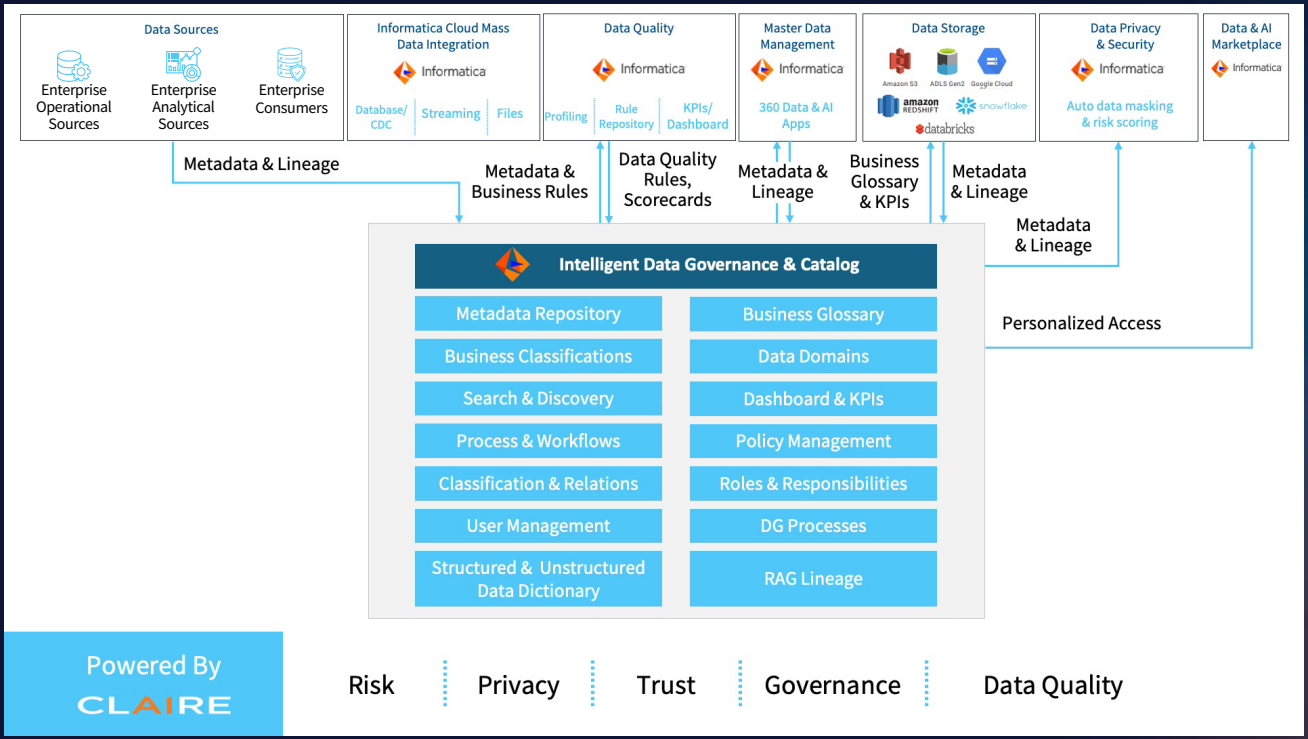


Figure 6: Intelligence Data Governance Framework Powered by Claire

The Informatica's AI-powered Intelligent Data Management Cloud (IDMC) platform focuses on modernizing data management and governance processes. Its "Data Governance, Access & Privacy," "Data Catalog," and "Data Quality & Observability" enable organizations to improve data trust and govern data across various environments such as on-prem and cloud. Informatica's AI engine, CLAIRE, enables the governance framework that not only enhances data management processes through intelligent automation but also improves efficiency by providing fast data insights. Some of the key features that make IDMC stand out, include:

- Advanced analytics intelligence locates all forms of data and delivers an enriched view, including its purpose, structure, and relationships. CLAIRE AI engine automates classifications of sensitive data, provides universal control to govern data access, and recommends relevant context and association.
- IDMC's Data Catalog understands the data ecosystem and explores the data lifecycle across the value chain by automatically creating a lineage.
- The Data Quality & Observability dashboard is integrated with IDMC's Data Governance, Access & Privacy, allowing users to view the overall trust score from a single panel.

- IDMC's Data Governance, Access, and Privacy provides complete visibility of data and AI models. It also automatically generates the business context of the metadata for a comprehensive view and accelerates the glossary documentation process.
- Informatica's Cloud Data Access Management (CDAM) is an AI-powered solution that automates data access by using the common metadata repository.

This way, we can see how data quality, data catalog, and data governance components of IDMC complement each other to achieve the end goal of maintaining data trust.

Strengthening data & AI governance with LTIMindtree

LTIMindtree helps organizations navigate through a combination of strategic consulting, cutting-edge technology, and tailored solutions that integrate AI into governance and security processes.

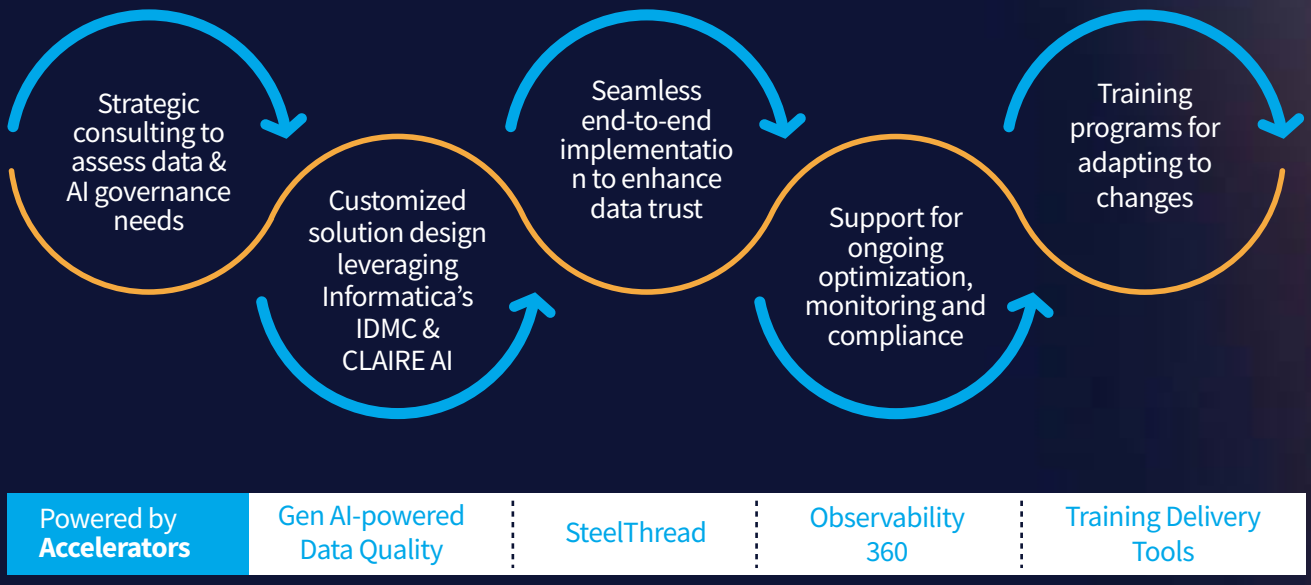


Figure 7: How LTIMindtree can help

1. **Strategic consulting:** LTIMindtree's Data Governance Consulting team evaluates the current data and AI governance framework across people, processes, and technologies. This helps us identify areas for enhancement and use cases where AI features can complement data trust and compliance. The goal is to develop a comprehensive strategy that aligns with the business goals and regulatory requirements.
2. **Customized solutions:** Leveraging the capabilities of Informatica's IDMC platform and with the support of LTIMindtree's Data Governance COE and Informatica COE, we design customized solutions that automate data management processes, data observability, compliance monitoring, and data governance operational workflows. Our expertise in the IDMC platform ensures that the data governance framework is not only efficient but also intelligent, providing actionable insights and real-time monitoring.
3. **End-to-end implementation:** To establish a strong data and AI governance framework, we ensure seamless execution of solution design, making sure all components work together seamlessly. We leverage the full potential of IDMC's AI features to maximize benefits. Additionally, the Informatica Partner Support program promptly addresses technical issues during implementation.
4. **Ongoing support and optimization:** We offer ongoing support to measure data governance efficiency and ROI, ensuring the framework remains effective and updated with regulatory changes and technological advancements.
5. **Training programs:** Our Data Governance COE and learning and development team ensure that our "People" are well-equipped to manage and utilize data effectively, keeping the data governance framework resilient and adaptive. We offer data governance organization's onboarding program, train-the-trainer sessions, and detailed documentation of Standard Operating Procedures (SOPs).

We provide Gen AI-augmented enablers to apply automation and speed up the journey. Here is a snippet view of our key accelerators:

- **Gen AI-powered data quality:** Lightweight Gen AI-based data profiling tool that can identify data types and key KPIs to recommend necessary Data Quality rules and identify anomalies.
- **Observability 360:** A metadata management and visualization tool that helps present quality metrics in different formats.
- **SteelThread:** Gen AI-powered lineage visualization utility to ensure effective mapping and business rule coverage during next-gen Master Data Management (MDM) implementation.
- **Training delivery tool:** Digital platform that enables users to walk through different functionalities of the process and application, offering step-by-step instructions to help them complete data governance and stewardship tasks.

LTIMindtree

Conclusion

As AI continues to shape the future of business, the importance of robust data governance cannot be overstated. Data Governance for AI ensures that the data used by AI systems is managed responsibly and ethically, while AI for Data Governance leverages AI technologies to enhance data governance practices. To stay ahead in the ever-evolving landscape of data governance and leverage AI's full potential, it's crucial to adopt robust tools and practices. IDMC's comprehensive suite of data governance solutions addresses both these aspects, enhances data security, achieves cloud data protection, and ensures AI data compliance.

References

1. Bill Gates explains why we shouldn't be afraid of A.I., Kif Leswing, CNBC, July 12, 2023: *https://www.cnbc.com/2023/07/12/bill-gates-explains-why-we-shouldnt-be-afraid-of-ai.html*
2. Data Quality Top Obstacle to GenAI, Informatica Survey Says, Alex Woodie, BIGDATAWIRE, February 5, 2024: *https://www.bigdatawire.com/2024/02/05/data-quality-top-obstacle-to-genai-informatica-survey-says/*
3. AI Governance Frameworks For Responsible AI, Gartner Peer Community: *https://www.gartner.com/peer-community/oneminuteinsights/omi-ai-governance-frameworks-responsible-ai-33q*

About the author

Ranjini Balakrishnan

Data Strategist, Data Architecture and Strategy, LTIMindtree

Ranjini has over 16 years of experience in data analytics strategy, data management, AI governance consulting, end-to-end implementation, and defining data literacy programs. She has assisted CXOs across various industries in establishing a vision, devising technical solutions, and charting a strategic road map for their data to decision journey.