LTIMindtree

Powering AI with the Right Data:

Building a Future-Ready Data Infrastructure

Abstract

AI is only as powerful as the data it learns from. Yet, many organizations struggle with fragmented, inconsistent, and unstructured data, limiting AI's full potential. This PoV explores the critical role of AI-ready data infrastructure—why it matters, the roadblocks enterprises face, and the key principles to get it right. From ingestion to governance, security, and scalability, we break down what it takes to build a robust data foundation that fuels AI-driven success. Whether you're optimizing existing systems or starting fresh, this paper offers actionable insights to transform your data strategy and accelerate AI adoption with confidence.

Introduction

As artificial intelligence (AI) continues to transform the IT landscape, one of the most pressing questions for enterprises is: where to begin? While challenges in adoption persist, ensuring AI-ready data remains a top concern for Chief Data Officers (CDOs) and Chief Information Officers (CIOs).

The phrase "garbage in, garbage out" has never been more relevant. AI models thrive on high-quality data, prompting organizations to refine their data infrastructure for better AI performance. However, it is not enough for data to exist in an ideal state within source systems—it must be relevant to business processes and readily consumable. This calls for a structured data management strategy, ensuring all sources are governed and harmonized to support an organization's AI journey.

Challenges organizations are facing in AI integration

A recent survey from Harvard Business Review confirms that business leaders recognize the importance of a reliable data foundation. While AI is widely seen as a major disruptor, data integration remains one of the primary obstacles to its adoption.

> 91% of survey respondents agree that an organization must have a reliable data foundation to successfully adopt artificial intelligence (AI)
>
> - Harvard Business Review

> 66% expect AI to be a major disruptor in their organization's industry within the next three years
>
> - Harvard Business Review

> 50% say that when it comes to data challenges their organization is experiencing in adopting AI, they're having difficulty integrating diverse data sources into a unified format
>
> - Harvard Business Review

> At the risk of sounding clichéd, data is gold. But data is also a commodity that ages quickly. So having the right data—quality data, accurate data, fresh data—is clearly a prerequisite for doing things that will give you a competitive advantage, says Florin Rotar, chief artificial intelligence (AI) officer at Avanade.
>
> - Sourced from Harvard Business Review

Figure: 1: Challenges organizations are facing to apply AI/Gen AI

Interestingly, organizations experimenting with AI models have encountered an unexpected challenge: AI's ability to surface information they were unaware of. This discovery, while insightful, also poses risks. Some companies prefer to keep certain data hidden, yet AI integration can inadvertently expose low-quality or sensitive data. “Gen AI can surface data you might not have been using. And if that data is bad, it can suddenly get exposed in a completely different way,” cautions Florin Rotar, Chief AI Officer at Avanade.

A 2024 Harvard Business Review study highlights the top challenges organizations face in adopting AI:

- Lack of necessary skills and expertise (55%)
- Poor **data quality due to siloes, duplication, or untrustworthy sources (48%)**

To address these concerns, organizations must take a strategic approach to preparing their data infrastructure for AI. In this POV, we will focus on how to make your current data platform ready.

Below are some of the other key challenges that organizations have reported in a recent study by Harvard Business Review.

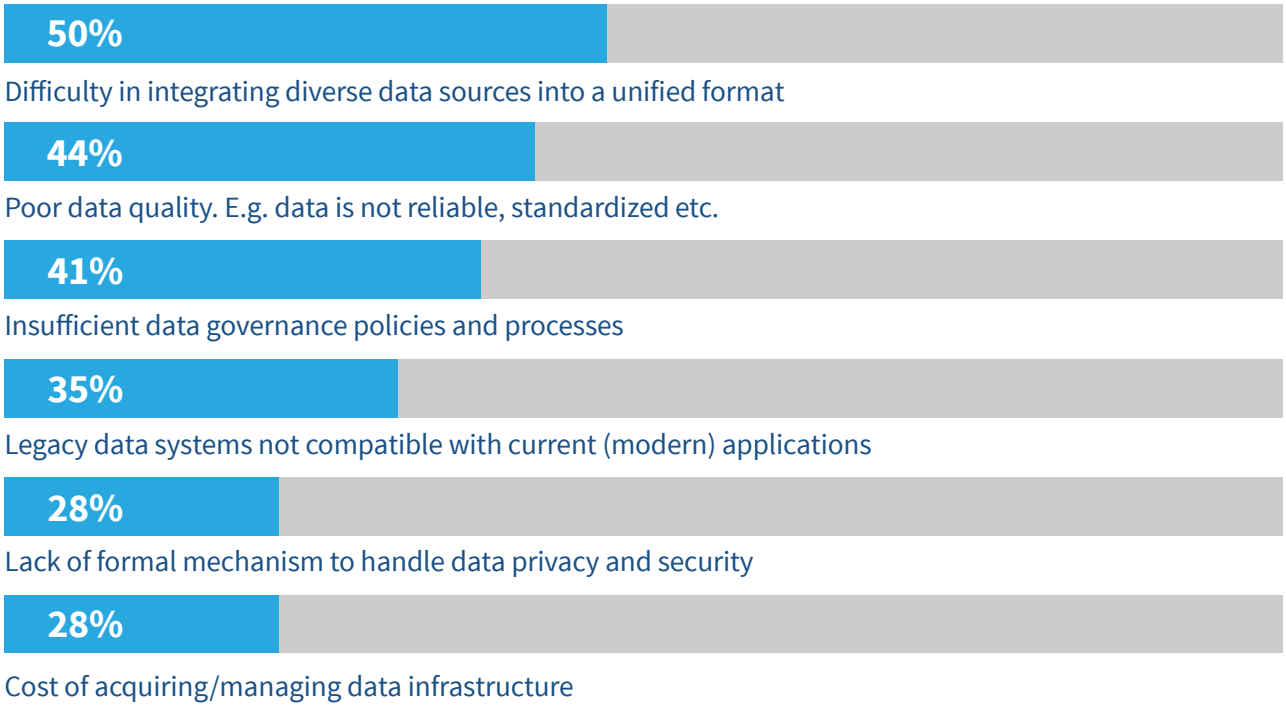


Figure: 2: Top challenges faced by organizations in adopting AI

Key data challenges hindering AI readiness

Several recurring issues prevent organizations from fully leveraging AI:

- **Limited understanding of source systems:**
  Many data platforms lack processes to assess and validate extracted data. This results in unused data clogging data infrastructure, increasing storage costs.
- **Misalignment with business processes:**
  Many organizations treat data platforms as separate IT initiatives, causing a disconnect between business needs and available data.
- **Fragmented and siloed data sources:**
  Dispersed data sources make integration difficult, reducing AI's effectiveness.
- **Poor data quality:**
  Poor data hygiene prevents organizations from building a reliable, consolidated data source for AI-driven analytics.
- **Scalability constraints:**
  AI models to function as per the specifications, need a lot of data to be fed in their respective models. The data landscapes in most of the organizations are typically built to meet the needs of their daily business operations. Here, the scalability dependencies are overlooked, bringing any AI-based development approach to a stop.
- **Inefficient data processing:**
  Whether real-time or batch processing, many data management strategies lack optimization frameworks tailored forAI-driven insights.

Building an AI-ready data platform

How to identify the starting point and prioritize the components which would need to go through the preparatory phases of data readiness? Now, let us focus on some of these areas which would help the data platforms become future ready to handle AI needs.

1. Aligning business use cases with AI objectives:

This is one of the most important areas to consider while building an optimum AI-based use case implementation. Understanding the relationship between AI use cases and business objectives is essential. Defining key use cases ensures leadership buy-in for data management strategies and establishes clear KPIs to measure return on investment (ROI).

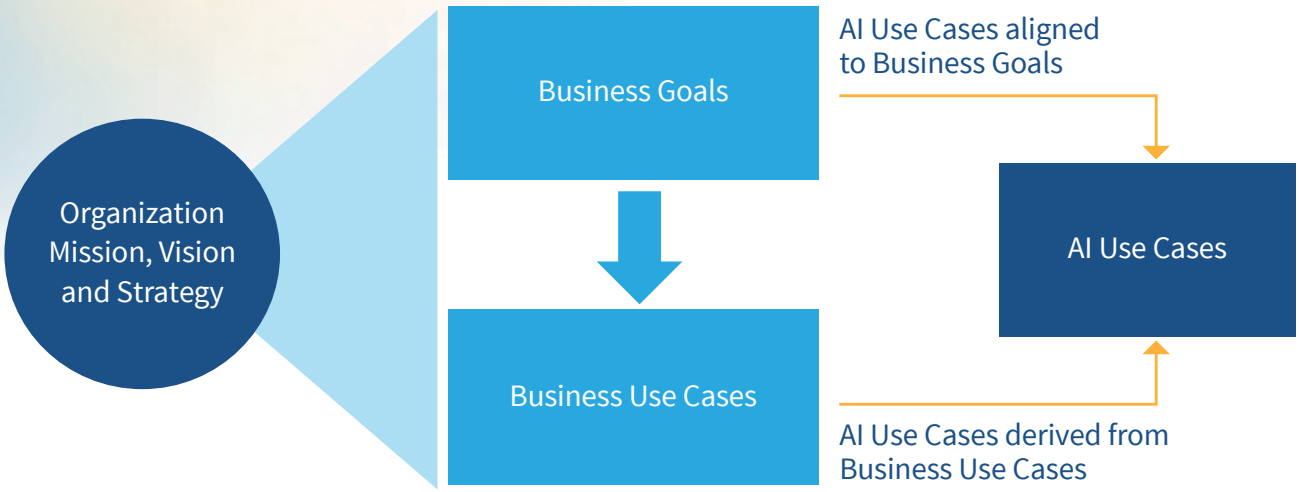


Figure: 3: Relationship between AI use cases and business use cases

2. **Establishing AI-specific design principles:**
   While building a data platform which would be relevant towards implementing AI-based use cases, specific design principles should also be considered. These design principles should be aligned to the AI maturity curve of the organization starting from "Ad hoc" to "stable" to "advanced" and "transformative". This progression reflects not just technological advancement but also cultural and strategic readiness. The parameters of this evolution must be aligned to the overall strategy, vision, and mission.

Some of the key design principles include:

- **Customer behavior adaptability:**
  Define principles which can predict what users can do or mostly do while they consume the data generated within your platform.
- **Enhanced customer engagement:**
  Customers should be able to connect with your offerings through the data you publish. Hence your AI model should be mature enough to provide insights that improve problem-solving and decision-making for end users.
- **Risk management and compliance:**
  One of the biggest barriers while developing AI-based applications is developing and following the process of security, risk, and compliance. In fact, in case a mistake occurs, it generates ripple effect towards affecting the organization's brand value.
- **Leveraging AI capabilities:**
  What are the capabilities of the AI process you are utilizing, and how you are gaining over human abilities? This should be part of your design strategy so that your design can use the generated data that is being produced, stored, and distributed across the organization.

Technical considerations for AI-ready data

To optimize data infrastructure for AI, organizations must refine various aspects of their data management strategies:

Data sources and ingestion

- Diverse data types (text, images, audio, video), format standardization challenges
- Scalability and modality-specific preprocessing needs
- Ensuring compatibility across modalities (e.g., text and image integration)
- Volumetric and velocity challenges of data, real time integration and IoT ingestion consistency
- Unstructured document, image and text content extraction, labelling and enrichment challenges
- Difficulty of semantic labelling and higher classification of data from sources

Data storage and processing

- Document, No-SQL and vector store efficiency
- Data virtualization, schema enforcement and data modelling challenges
- Synchronized processing of multimodal data
- Normalization and feather extraction
- Computational complexity and scaling massive amounts of unstructured data

Data quality, governance and lineage

- Unstructured data can have multiple classification tags and missing context
- Handling inconsistent and noisy data, data bias at scale
- Provisioning high-quality data to fine tune generative data models can be a challenge
- Data provenance and traceability

Data security and privacy

- Data privacy and compliance with regulations challenges
- Handling sensitive PII data securely, anonymization of data
- Implementing access control centrally or in a federated manner
- Responsible AI framework implementation
- Row and column level security principles at different layers of featured data set can be a challenge
- Robust role-based access control in data exchange and sharing endpoints

AI/ML model operations

- Fine-tuning models with domain-specific data
- Balancing model accuracy and computational efficiency
- Large-scale, multimodal training datasets
- Ethical usage, prevention of data leakage when using LLMs

Model deployment and inferencing

- Monitoring model performance and managing drift
- Ensuring efficient inference across modalities
- Ensuring robust ML-Ops practices
- Handling deployment of multimodal models

Analytics and consumption

- Accessible and interpretable analytics for a broad spectrum of users
- Meaningful generative AI apps that are not able to substantiate facts relevant to the use cases
- Integration with Data Marketplace apps with scaling of users and usage hours
- Self-serve data analytics platforms integration and scalability
- Scaling of analytics platform

Establishing a strong AI/GenAI foundation for data readiness

While building a data foundation for your AI-specific use case, what becomes crucial is how the data architecture or platform can offer trust among the consuming processes. The scalability parameters should be set up in such a way that it brings robustness and addresses the security aspects of the information. The data architecture design should include these aspects of data handling and include the relevant technological infusions.

While there has been substantial differentiation in the type of data that needs to be processed for accuracy in the AI-led processes, the storage and compute mechanisms have also responded to this cause. The data preparation process has been more structured, and these measures should be included while designing the data platform. Security, risk and compliance-related use cases should be given priority and the LLMs should be fed with data which restricts biasness and behaves responsibly.

Figure: 4: Foundation for AI/GenAI data readiness

Enabling AI readiness with LTIMindtree and Informatica

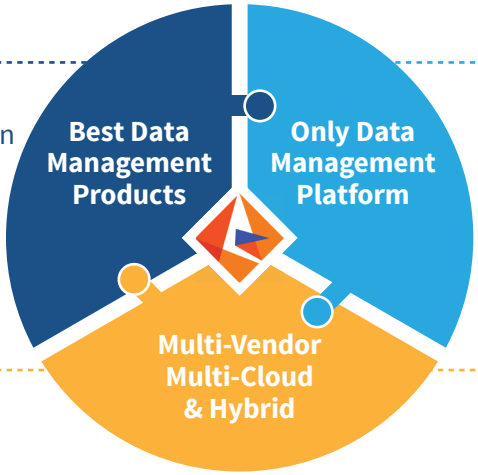
Becoming AI-ready starts with building a strong data foundation. Informatica, a leader in data management, provides comprehensive solutions for data governance, quality, and integration. Combined with LTIMindtree's expertise, organizations can build scalable, AI-ready platforms that drive efficiency, innovation, and long-term business value.

Here's how Informatica can help organizations become AI-ready:

Informatica Today

The Data Management Choice for Enterprise AI

- Born in the cloud to support AI/GenAI journey by any modern data architecture patterns
- 3rd-party validated leadership in every category
- AI-enabled products in production today



- Only true platform for data management
- Only AI-powered data management platform built on metadata system of record
- IPU consumption pricing model

- Broadest set of connectors, accelerators, and multi-cloud integration tools
- "Switzerland of data"

- Manage data across any cloud platforms, on-premises, and hybrid environments
- Supports any AI platforms

Figure: 5: Enabling AI-ready data management with Informatica

GenAI Use Cases

How Informatica helps

Use Cases	Prompt Engineering	Retrieval-Augmented Generation (RAG)	AI Agent(s) Systems	Fine Tuning
Description	• The art and science of crafting clear instructions (prompts) for AI language models. • Enables generation of accurate and helpful responses from AI models.	• Incorporating authoritative or proprietary data sources into GenAI Models. • Improves accuracy, trustworthiness, and context of LLM responses.	• An AI program capable of autonomous decision-making and acting on behalf of users, systems, or other programs.	• Tailoring GenAI models through focused training on specific datasets to enhance their performance in tackling particular tasks.
Informatica Enables	Simplified, low-code, orchestrated and governed Prompt Engineering	Automated, no/low-code pipelines with trusted knowledge	Scalable, no/low-code, governed AI Agent(s) trained on authoritative data	High-quality data to tune GenAI models through cleansing, API management & governance
Informatica Services for GenAI	• API & APP integration • API management • Out-of-the-box connectors • Prebuilt Prompt Recipes • Governance & Privacy for AI	• MDM • Data Quality • API & APP integration • API management • Data Engineering & Integration • Prebuilt Recipes for RAG pipelines • Governance & Privacy for AI	• API & APP integration • API management • Data Engineering & Integration • Prebuilt Recipes for Agent pipelines • Governance & Privacy for AI	• Data Quality • Data Engineering & Integration • Prebuilt integration pipeline templates • Headless data management, INFACore • API & App Integration • Governance & Privacy for AI

Figure: 6 Key Gen AI use cases supported by Informatica

LTIMindtree, a strategic partner of Informatica, brings deep expertise in large-scale data readiness programs. Our LTIMindtree AI platform accelerates AI adoption with industry-specific solutions (like regulatory and compliance lineage, and anomaly detection). By integrating Informatica's metadata-driven data management with our AI-powered solutions, organizations can establish a scalable, trusted data foundation. Our Gen AI-powered data governance reference architecture provides a structured approach to aligning enterprise strategies with implementation, supporting modernization, re-platforming, and greenfield initiatives. Together, Informatica and LTIMindtree help organizations achieve data maturity, ensuring their AI initiatives are built on a solid, future-ready foundation

Conclusion

As organizations rush to embrace AI and Gen AI-driven operations, data custodians must ensure they bring the right data into their platforms for consumption. The key data management aspects, like data quality and data governance are no longer optional—they are essential. Platforms like Informatica provide end-to-end data management capabilities, addressing the core data needs within an organization and ensuring AI initiatives are built on a reliable, high-quality foundation.

Author

Sumit Mukherjee
Data Solution Architect, LTIMindtree

Sumit is a Data Solution Architect in the Data Management Center of Excellence, bringing over 24 years of experience in designing and implementing data solutions. He has played a key role in guiding organizations through their data transformation journeys, helping them build scalable, efficient, and future-ready data ecosystems.

References

- Getting Your Data Ready for AI, Naveen Patlola, Informatica, September 17, 2024: *https://www.informatica.com/content/dam/informatica-cxp/techtuesdays-slides-pdf/Getting%20Your%20Data%20Ready%20for%20AI.pdf*
- Is Your Company's Data Ready for Generative AI?, Thomas H. Davenport and Priyanka Tiwari, Harvard Business Review, March 26, 2024: *https://hbr.org/2024/03/is-your-companys-data-ready-for-generative-ai*

LTIMindtree is a global technology consulting and digital solutions company that enables enterprises across industries to reimagine business models, accelerate innovation, and maximize growth by harnessing digital technologies. As a digital transformation partner to more than 700 clients, LTIMindtree brings extensive domain and technology expertise to help drive superior competitive differentiation, customer experiences, and business outcomes in a converging world. Powered by 86,000+ talented and entrepreneurial professionals across more than 40 countries, LTIMindtree — a Larsen & Toubro Group company — solves the most complex business challenges and delivers transformation at scale.
For more information, please visit **https://www.ltimindtree.com/.**