SPOTLIGHT

Sponsored by: Tata Communications

An AI-ready data platform serves as the core of the data ecosystem and must deliver unified management, cloud-edge flexibility, support diverse data sources and guarantee robust security with comprehensive data governance.

AI-Ready Data: Your Foundation for AI-Fueled Business

[March 2025]

Written by: Surjyadeb Goswami, Research Director

Introduction

AT A GLANCE

KEY STATS

- 20% AI projects failed in Asia/Pacific, mostly due to data related challenges
- In Asia/Pacific, by 2028, 161 ZB of data will be generated across multiple form and environments, growing at 24% CAGR since 2023

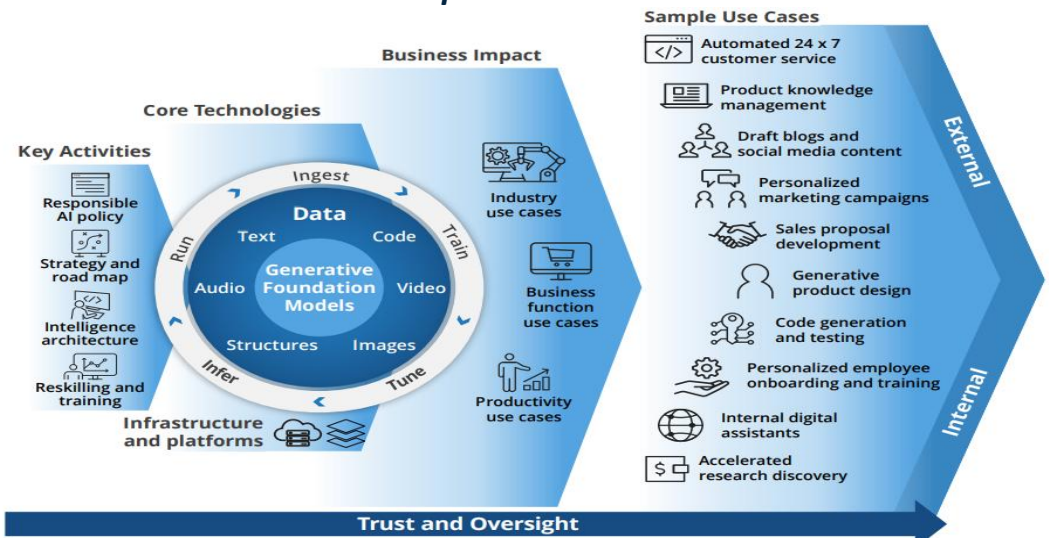
Artificial Intelligence (AI) and Machine Learning (ML) adoption is a critical step in evolving into an AI-fueled business. An early adoption of AI helps laying the groundwork for a robust data infrastructure, advanced analytics, and the right human capital development. As the adoption matures, AI technologies enhance organizational competitiveness and drive operational excellence. As per IDC's July 2024 *Asia/Pacific Data, AI, GenAI and Insights Survey* (N=450), over 80% of organizations in the Asia/Pacific region are already using AI to optimize operations, enhance decision-making, and unlock insights from vast data. As an ongoing effort, many organizations are now adopting Generative AI (GenAI) technologies as it offers new opportunities for content generation, personalized experiences, and automating complex tasks, further driving innovation and business value across the region. According to IDC's December 2024 *Future Enterprise Resilience and Spending Survey, Wave 11,* (n=300 for Asia/Pacific), about 48% of organizations in Asia/Pacific have already introduced several GenAI-enhanced applications and services into production, while 37% plan to launch in the next 12 months. By embracing GenAI across their operations, organizations in the region are not only enhancing their operational capabilities but are also laying the foundation for long-term success in their digital transformation journey.

Scaling AI initiatives for impactful business outcomes, however, remain a significant hurdle. There has been an enormous level of data growth across multiple data environments and that adds complexity to the implementation journey. As per IDC's May 2024 *Global Data Sphere*, in Asia/Pacific region, 161 ZB of data will be generated by 2028, growing at 24% CAGR since 2023, the growth is even higher at 33% for enterprise data at the Edge. In key countries like India, though AI adoption is accelerating, challenges persist, particularly around data quality issues, noise, and failures, moreover, according to IDC's July 2024, *Asia/Pacific Data, AI, GenAI and Insights Survey* (n=50 for India), around 28% of organizations call out data trustworthiness/bias key reason that hinder successful AI implementation. To fully unlock data-driven business value,

organizations must prioritize trustworthy, high-quality data as the cornerstone of their AI strategy. This will also depend on how organizations implement AI technologies across various use cases.

IDC has developed the GenAI path to impact (see Figure 1) and provides a roadmap for understanding the prerequisites to optimize AI for the desired business outcome

FIGURE 1: ***IDC's GenAI Path to Impact***



Source: IDC

Why AI Projects Fail

According to IDC's July 2024 *Asia/Pacific Data, AI, GenAI and Insights Survey* (n=50 for India), 20% of AI projects in Asia/Pacific have failed. This can be attibuted to factors such as governance and risk management issues, skill shortage, high infrastructure costs, and definitely the challenges around "Data", from the quality to the management of it.

Reasons for failure:

- **Data challenges:** Challenges like restricted data access, dynamic or expiring datasets, lack of adequate volume of properly trained datasets, biased or poor-quality data, and business-imposed restrictions, further intensified by the complexities of data engineering.
- **Governance and Security challenges:** Weak governance, poor data quality, bias, and misaligned strategies hinder AI adoption, requiring robust frameworks to keep pace with rapid technological change. Data security around unauthorized access, breaches and manipulation also becomes crucial for AI projects to be successful.
- **Skills gap:** A lack of AI expertise and data literacy limits adoption, necessitating investments in upskilling and talent acquisition to drive AI-led innovation.
- **Cost barriers:** High costs for talent, infrastructure, and data management, combined with technology immaturity and resource disparities, challenge AI adoption, demanding smarter cost strategies and resource optimization.

FIGURE 2: ***Reasons for Failure of AI Projects***

Q. Aggregated - Please select the top 3 challenges involved in implementing AI technology in your organization.

Source: IDC's Asia/Pacific Data, AI, GenAI and Insights Survey, July 2024, Asia/Pacific N=450

Success demands robust data strategies, ensuring accessibility, accuracy, bias mitigation, and governance to power reliable AI outcomes.

Drive Towards a Successful GenAI Journey: The Critical Role of a Strong Data Value Chain

To scale the benefits of GenAI initiatives, organizations must understand the concept of a data value chain to ensure that they can leverage the immense potential of GenAI but do so with the necessary data governance and control.

As a strategic imperative, the GenAI data value chain is the backbone of successful GenAI initiatives. It orchestrates the sourcing, processing, and management of diverse, high-quality data that fuels foundational and fine-tuned AI models. By establishing a strong data foundation, the value chain ensures that GenAI delivers accurate, safe, and meaningful outputs while empowering organizations to innovate responsibly.

Building a robust GenAI data value chain: key processes, functions, platform and stakeholders

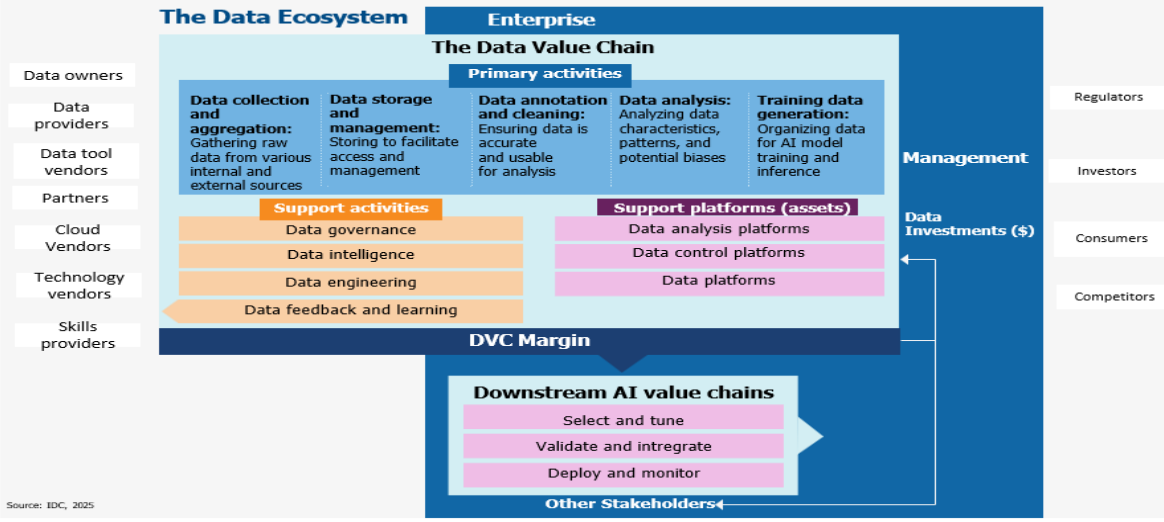
The GenAI data value chain encapsulates the entire life cycle of data management processes that transform raw data into insightful, valuable outcomes in GenAI contexts. Building a strong GenAI data pipeline as part of the primary stages of the

data value chain involves collecting and managing diverse, high-quality data, whether structured or unstructured, for training and inference. Storage and cleansing ensure performance and relevance, while annotation and analysis enhance contextual accuracy. Training data preparation refines datasets for AI models, and robust data protection safeguards sensitive information against exposure.

Supporting these primary activities are secondary activities such as data governance, data intelligence, data engineering, and data feedback and learning that cut across all stages of the data value chain.

Many infrastructure assets are used in both primary and secondary activities. Starting from the most foundational, the data itself, these include data platform, data control platform, that helps manage, secure and control data usage and access across the enterprise. As the data analysis platform transforms raw data into actionable insights, it becomes an invaluable asset, particularly for developers.

FIGURE 3: ***The Data Value Chain and Its Supporting Ecosystem***



Source: IDC PlanScape: GenAI Data Value Chain; IDC #US51946724

There are multiple internal stakeholders of this GenAI data value chain – CISOs in GenAI safeguards data integrity and enforce security policies. Their focus on preventing data breaches and establishing a secure AI infrastructure is critical to maintaining the integrity and trustworthiness of AI systems, applications, and infrastructures. There are also CDOs who oversee data acquisition, quality and ensure alignment with organizational goals. There are data engineers, data scientists, ensuring design, maintenance of data infrastructure and developing or refining GenAI models. The list is not limited to the above only and it is essential for the teams to work in collaboration to gain real value from the GenAI data value chain.

Key considerations for enterprises while building a GenAI data value chain

- **Strategic data acquisition and ecosystem partnerships**: Organizations need to look for the right data sources and partnership which ensures high quality and domain specific data to run AI models. It is also recommended to build a sustainable data monetization strategy and encourage respective teams/ departments, which will drive a continuous flow of reliable data. A strong data acquisition strategy ensures AI systems are trained on diverse, unbiased, and relevant datasets.

- **Ensuring data quality through cleansing, validation, and structuring:** Automating the process of data ingestion, validation, and enrichment enhances model accuracy and efficiency. Creating a robust data architecture supports multimodal data structures, ensuring adaptability across AI use cases. Vector databases play a crucial role in efficiently storing embeddings, enabling semantic search, personalized recommendations, and real-time AI applications requiring rapid data retrieval.
- **Optimizing data processing for AI performance:** Techniques like tokenization and chunking enable AI models to efficiently process and interpret large datasets. Tokenization enhances language understanding by breaking text into meaningful units, while chunking improves scalability by enabling parallel processing. These indeed take care of memory overhead, drives optimization of compute resources, and ensure seamless AI-driven insights across various applications.

In the GenAI race, the data value chain is a strategic enabler for innovation, resilience, and competitive advantage.

- **Deliver trustworthy and high-impact AI outputs:** A robust data value chain transforms raw data into a trusted asset, ensuring that your AI outputs are not only accurate but also safe, helping to avoid reputational or operational risks.
- **Scale AI with confidence and control:** The value chain provides a clear, governed framework for managing the complexities of data, enabling your organization to scale GenAI initiatives without compromising on compliance, quality, or control.
- **Unlock new levels of business efficiency:** By systematically adding value to data at every stage, from acquisition to actionable insights, the GenAI data value chain empowers your organization to innovate faster, optimize operations, and achieve meaningful business outcomes.

Deliver AI-ready Data in the Intelligent Enterprise

Organizations often invest in collecting and processing data, but the key to making this data useful for AI is managing it effectively. This means going beyond basic data management, which ensures data is accurate and consistent, to treating data as a valuable product.

By doing so, organizations can better understand how to make data easy to use and manage, ultimately unlocking its value for AI applications. Therefore, businesses need to consider different aspects of data management, that starts with identifying the data locations, sources, data types, data changes and events, and enabling access via open data APIs. Following which, implementing policy and control for data access, engineering, quality, and data sharing becomes crucial. The full value of data is realized when it is used in business activities, delivering required business outcomes that are driven and informed by data and to arrive there, data synthesis activities include business analytics, advanced analytics, machine learning, and AI model tuning, training, and validation activities need to be executed on the available data.

Many organizations address data-driven decision-making through isolated projects and investments in AI tools, and governance but fail to integrate technologies, processes, and initiatives to align with the broader goal of enhancing enterprise intelligence. That is where the real value of the GenAI data value chain diminishes. An organization needs to understand the different layers of data architecture in detail, analyze the challenges of implementing the architecture and evaluate ways on how intelligent automation can drive improved data quality, delivery of insights at scale, enabling data literacy and promoting data culture.

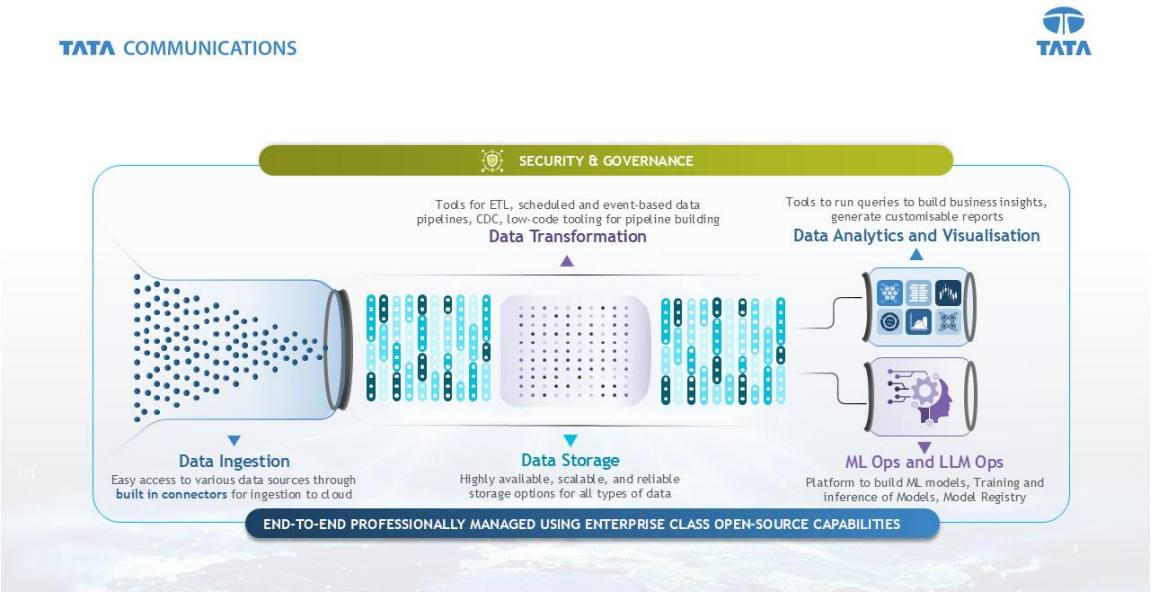
An enterprise intelligence architecture is critical to support continuous and collective learning, delivery of insights at scale, enabling data literacy and promoting data culture. This consists of below four planes, and this is paramount for organizations to focus their AI-ready data journey through the lens of this intelligent architecture.

- **Data plane:** Organizations manage diverse, distributed, and dynamic data using databases, data warehouses, and datalakes for structured and unstructured formats
- **Data control plane:** Data is engineered to support different business activities. It is where data engineers put context around the data and where governance such as data integrity is applied to support AI models.
- **Data synthesis plane:** Domain-specific data synthesized into business-focused data products through advanced analytics and AI.
- **Business activity plane:** Business actions such as decisioning, optimization, publication, cataloguing, communication take place as a result of applying business specific problems for trusted outcomes

Considering Tata Communications Vayu Data Platform

Tata Communications Vayu Data Platform is a unified, cloud-to-edge-agnostic solution designed for seamless data management, analytics, and AI workloads. It enables organizations to process big data, generate insights, and manage machine learning workflows with end-to-end lifecycle management—from data ingestion to AI-driven decision-making. Built on open-source technologies with no vendor lock-in, it reduces TCO and future-proofs investments by continuously evolving with new capabilities thus helping in creating the right data ecosystem partnerships. Its flexible architecture supports structured, semi-structured, and unstructured data, ensuring effortless integration with existing tools and optimizing data processing for improved AI performance. With advanced MLOps automation, self-provisioning, and connected intelligence, the platform accelerates AI adoption while enhancing efficiency, security, and scalability and delivering the real business value of AI.

FIGURE 4: ***Tata Communications Vayu Data Platform***



Source: Tata Communications

With the proposition of "secure by design" the Vayu Data platform offers advanced features like fine-grained access control, data masking, consent management, and PII protection to safeguard sensitive organization data. The platform also takes care of data governance and compliance requirement for the organization, through metadata management classification, and lineage tracking, adhering to privacy regulations such as PCI DSS, CERT-In, and HIPAA, that helps organizations to meet security and compliance requirements necessary for being AI-ready data.

Challenges

- **Brand recognition:** Closely related to both competition and scale, larger competitors may have a stronger brand presence and deeper client relationships when it comes to Data products. Building brand awareness and establishing credibility with potential clients will require sustained investment by Tata Communications in marketing and business development activities.
- **Strong Partner Ecosystem:** While partnerships are essential, the company needs to carefully curate and manage its partner ecosystem to ensure a complete offering across the data layers, alignment with its strategic goals and deliver value to customers.
- **Differentiating its Offerings:** In a crowded market, Tata Communications needs to clearly differentiate its offerings by focusing on specific strengths, such as its global network, industry expertise, managed services capability and focus on emerging markets.

Despite these challenges, Tata Communications has demonstrated a strong commitment to innovation, customer success, and operational excellence, positioning itself as a strong provider of Data platform ensuring a success journey for organization's AI transformation.

Conclusion

To make the most of AI's transformative power, organizations need an enterprise-wide strategy for data that fits their business goals and builds a data driven culture. Business leaders must champion data driven decision making, ensuring analytics and insights guide innovation across diverse domains of organization.

A reliable -managed data platform plays a crucial role in maximizing AI's capabilities and sparking lasting business outcome

A scalable data foundation is critical for AI's success. Organizations should invest in cloud-native data architectures that enable flexibility and adaptability for future AI advancements. Ensuring scalability across the technology stack including infrastructure, data platforms, and foundational models allows businesses to expand AI applications beyond initial use cases to successful commercial deployments. Modular, open-standard architectures future-proof AI investments, enabling seamless integration with evolving technology landscape.

Creating a strong data value chain is crucial to fuel AI success. Collaborating between departments, suppliers, vendors, and customers ensures efficient data sourcing, engineering, security, governance and data management. A well-structured data ecosystem not just drives AI innovation but helps in deriving the real business value out of the investment made and empowering the organization to lead in the AI-driven future.

About the Analyst

Surjyadeb Goswami, *Research Director*

Surjyadeb Goswami leads research programs around AI and automation, focusing on AI and GenAI (GenAI) strategies and technology landscape along with end-user insights. In a previous stint with IDC, Surjyadeb led the Asia/Pacific Enterprise Storage and Networking Infrastructure research practice over a decade.

MESSAGE FROM THE SPONSOR

Fragmented and ungoverned data slows decision-making, increases risks, and drives up costs. The Tata Communications Vayu Data Platform provides an adaptive framework that unifies the entire data lifecycle—from ingestion to insight. With built-in governance and compliance, it simplifies data ownership, privacy, and regulatory adherence—no matter where your data resides. AI Ready and designed for scale, the platform transforms raw data into actionable intelligence—eliminating inefficiencies and unlocking business value. Enterprises can harness data to drive competitive differentiation, operational resilience, and sustainable growth.

IDC Custom Solutions

The content in this paper was adapted from existing IDC research published on www.idc.com.

IDC Research, Inc.
140 Kendrick Street
Building B
Needham, MA 02494, USA
T 508.872.8200
F 508.935.4015
Twitter @IDC
blogs.idc.com
www.idc.com