IMPORTANCE OF DATA FOR AN AI ASPIRING ORGANIZATION

Infosys®
Navigate your next
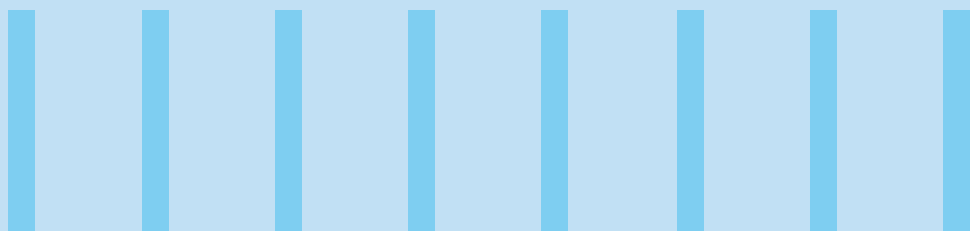
The relationship between AI (artificial intelligence) and data is fundamentally symbiotic and inseparable. Data serves as the cornerstone of AI development, right from initial model training and dataset curation to validation and performance optimization. Once deployed, AI models enable organizations to generate actionable insights and data-driven predictions that directly impact business outcomes and unlock significant value across operational domains.

However, the path from raw data to transformative AI capabilities for any organization presents considerable complexities and challenges. With the understanding that data underpins any AI capability, IT organizations must develop a comprehensive understanding of their data landscape through the strategic lens of key business drivers. This perspective is essential for organizations seeking to successfully navigate their AI transformation journey and realize sustainable competitive advantages in an increasingly data-driven marketplace.

The Evolution of Enterprise Data - Key Trends

Contemporary enterprises operate within increasingly complex data ecosystems that span entire supply chain networks, encompassing suppliers, partners, customers, and internal operations. The scope of data evolved from monolithic applications that held all the relevant data within themselves to big data warehouses, operational data stores and advanced Extract, Transform, and Load (ETL) technologies that aggregate data from heterogeneous sources, apply necessary transformations and consolidate information into unified repositories that support complex analytical workloads.

Where the data world is moving towards

Generative AI and Automated Insights	Generative AI tools are revolutionizing data analysis by enabling natural language queries. AI-powered analytics are becoming more sophisticated, with capabilities to detect subtle patterns and automate data preparation and insight quality.
Generative AI and Automated Insights	Generative AI tools are revolutionizing data analysis by enabling natural language queries. AI-powered analytics are becoming more sophisticated, with capabilities to detect subtle patterns and automate data preparation and insight quality.
Synthetic Data	Computer-generated data to mimic real data without compromising privacy, is gaining prominence as a solution to stringent data privacy regulations and ethical concerns.
Enhanced Data Governance and Security	Zero trust security models, emphasizing continuous verification rather than implicit trust, are widely adopted to protect sensitive data. Data governance gains importance amid complex regulatory environments.
Data Lakehouse & Open Table Formats	Data Lakehouse combines data lakes' flexibility with data warehouses' performance to dominate big data architectures. Open table formats enable standardized, cross-platform management of large-scale tabular data, reducing vendor lock-in and supporting evolving data structures.
Personalized Customer Experiences	Hyper-personalization based on customer behavior data uses augmented analytics to automate data discovery and interpretation is essential for engagement and retention for non-technical users.

Challenges Confronting Chief Data Officers (CDO) in Modern Enterprises

Top challenges faced by Chief Data Officers

Data Trustworthiness

- With data flowing in from multiple platforms like cloud services, IoT, third-party APIs, ensuring consistency and integrity becomes increasingly complex.

- Missing or partial data due to system limitations, integration failures, or lack of human oversight can skew analysis

- Errors from manual entry, outdated systems, and poor validation processes impacts the data trustworthiness.

Data Quality and Governance

- Data is often scattered across multiple systems and departments, making it an organizational hurdle for further processing.

- Poor data quality and inconsistent data readiness slow down the delivery of data products.

- Absence of a unified data management strategy and governance framework allows departments to independently select and implement tools without any formal business case.

Organization readiness

- Cultural resistance to data-driven approaches and low data literacy across the enterprise hamper effective data utilization.

- Ensuring data is trustworthy, accessible, and business-ready remains a persistent challenge for data teams.

- Centralized span of control/ rigid organizational structures and lack of leadership support is also a deterrent for the data teams.

Data Protection Measures

- Rising Threats and High Costs: High-profile data breaches impacting firms. CDOs must protect sensitive data to avoid such costly incidents.

- Regulatory Pressure: Compliance with regulations like GDPR and CCPA requires CDOs to maintain a full inventory of organizational data, understand its flow, and ensure it is stored as per following compliance requirements.

Data ecosystem/ platform

- AI needs clean and structured data and without a proper data platform, deploying intelligent automation is difficult.

- New tools are often added reactively to solve immediate problems, without evaluating existing capabilities or long-term integration.

- Rise of SaaS, cloud platforms, IoT etc. has increased the variety and volume of data inputs, further complicating the technology landscape.

Talent Retention

- The current workforce possess legacy skills but lack motivation to nurture new skills and technologies. This leads to a limited capability to process and derive insights from data.

- Limited set of data engineers and stewards operate beyond their capacity , thereby restricting their ability to innovate and develop new solutions.

Despite the recognized importance of data, organizations continue to face persistent data-related challenges that impede their AI readiness. Critical gaps exist across the entire data lifecycle, spanning from data collection and ingestion, pipeline development and transformation workflows, storage management to governance, modelling frameworks, data aggregation and labelling processes, and fundamental analytics capabilities. These deficiencies are compounded by widespread non-compliance with regulatory frameworks such as the General Data Protection Regulation (GDPR) and the Digital Personal Data Protection Act (DPDPA), creating additional operational and legal risks.

The absence of appropriate technology to effectively process data and extract meaningful insights represents a significant barrier to informed decision-making. This technological gap often results in compromised data quality regarding customer and prospect information, directly undermining organizations' ability to execute targeted engagement strategies and develop robust sales pipelines. Consequently, these systemic data challenges create a cascading effect that diminishes overall business performance and competitive positioning.

Five V's of enabling a data-driven approach

Variety

In today's data-driven landscape, data variety allows organizations to harness the full spectrum of available information-from structured records to unstructured social media and sensor data-thus unlocking richer insights, driving innovation, and enhanced decision-making. Managing and integrating this variety effectively through advanced technologies, governance, and data ecosystems is a strategic imperative for business success in 2025 and beyond. The richness of varied data enables companies to monetize data assets through Data-as-a-Service (DaaS) offerings and data marketplaces, creating additional revenue streams and fostering collaboration across industries and is essential for training robust AI and machine learning models. Some of the data types are as follows:

Structured Data
Organized and stored in databases with clear relationships. E.g. Customer information, financial transactions, inventory records.

Unstructured Data
Data that lacks a predefined format, requiring specialized processing. E.g. Emails, social media posts like Twitter and LinkedIn posts, videos, audio recordings.

Semi-structured Data
Data contains both structured and unstructured elements. E.g. XML, JSON files, log files, sensor data.

Real-time Data
Continuously generated and processed instantly. E.g.: IoT device outputs, stock market fluctuations, live traffic updates.

Big Data
Large-scale data that requires powerful computing. E.g.: E-commerce transaction logs.

Geospatial Data
These are related to geographic locations and spatial relationships. E.g. GPS coordinates, satellite imagery, maps.

Machine-generated Data
These are produced by automated systems and devices. RFID scans, server logs, automation system outputs would be a few examples of these.

Post classification of data, one can combine data using data integration tools and create a unified view of data across platforms. Consistency of data can be achieved through data normalization. Use of metadata and tagging along with usage of text and visual mining would help you to derive meaningful insights from data. It is also imperative to regularly monitor these data pipelines to avoid compromising the reliability of the system.

Velocity

Data velocity represents the rate at which information is generated, transmitted, and processed within an organization's ecosystem. In today's digital landscape, organizational competitiveness is increasingly determined by the ability to capture, analyse, and act upon data streams in real-time or near real-time. High-velocity data environments enable organizations to transform raw information into actionable insights instantaneously, thereby facilitating agile decision-making and responsive operations.

Modern velocity-oriented architectures leverage specialized databases and streaming analytics platforms to process continuous data flows without traditional batch processing limitations. These systems are particularly critical for handling high-frequency data sources such as Internet of Things (IoT) sensor networks, social media streams, financial market feeds, and telemetry data.

It helps an organization to transform raw, rapidly generated data into actionable insights in real time. This capability empowers organizations to be more responsive, customer-centric, and innovative, while improving operational efficiency and risk management. As data generation accelerates, maintaining high data velocity is a strategic imperative for businesses aiming to thrive in today's fast-paced digital environment.

Improving velocity can lead to faster buying decisions, agility in product development, and positive bottom line. For customer-oriented organization, accumulating customer information quickly and accurately helps them in understanding the customer challenges, anticipating their future needs and as a result providing them with personalized experiences through timely interactions.

Veracity

This parameter refers to the reliability and trust quotient of data and in current scenario, veracity of data is fundamental for enterprises to squeeze the real value from its data. It directly influences the quality of insights, the effectiveness of decisions, operational success, and risk management. Organizations that emphasize data veracity through strong governance, validation, and management practices are better positioned to leverage data for competitive advantage and innovation.

Ensuring data veracity is crucial for making informed decisions, as unreliable data can lead to flawed insights and poor business outcomes. Data inconsistencies, missing values, and biases are common in large data sets.

Prioritizing veracity involves:

Data source validation

Single unified view of data across different systems

Data cleansing and governance to convert raw data into reliable information that can be used for analysis.
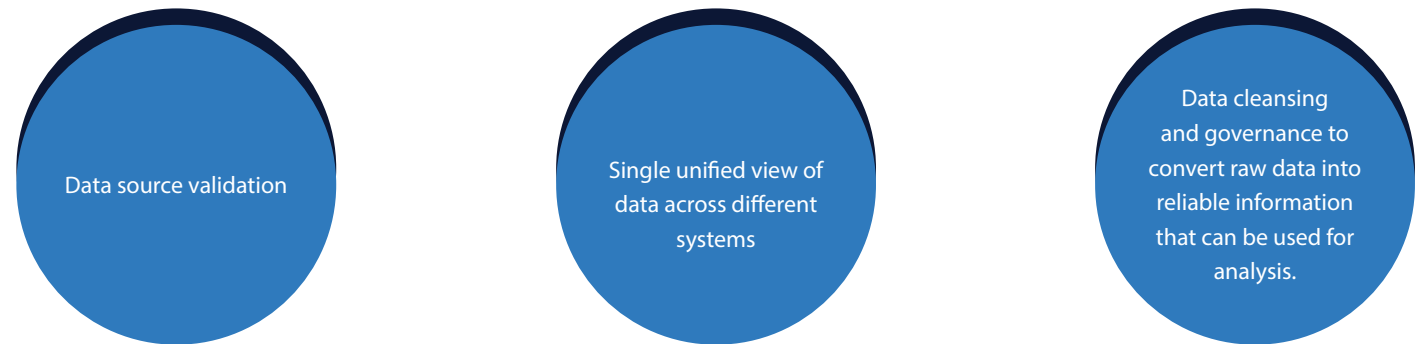
High data veracity builds confidence in analytics and reporting, and it also supports data security by ensuring only legitimate, accurate data enter systems, reducing manipulation and breaches.

Prioritizing data veracity not only boosts informed decision-making but also enhances precision and trustworthiness of data.

Value

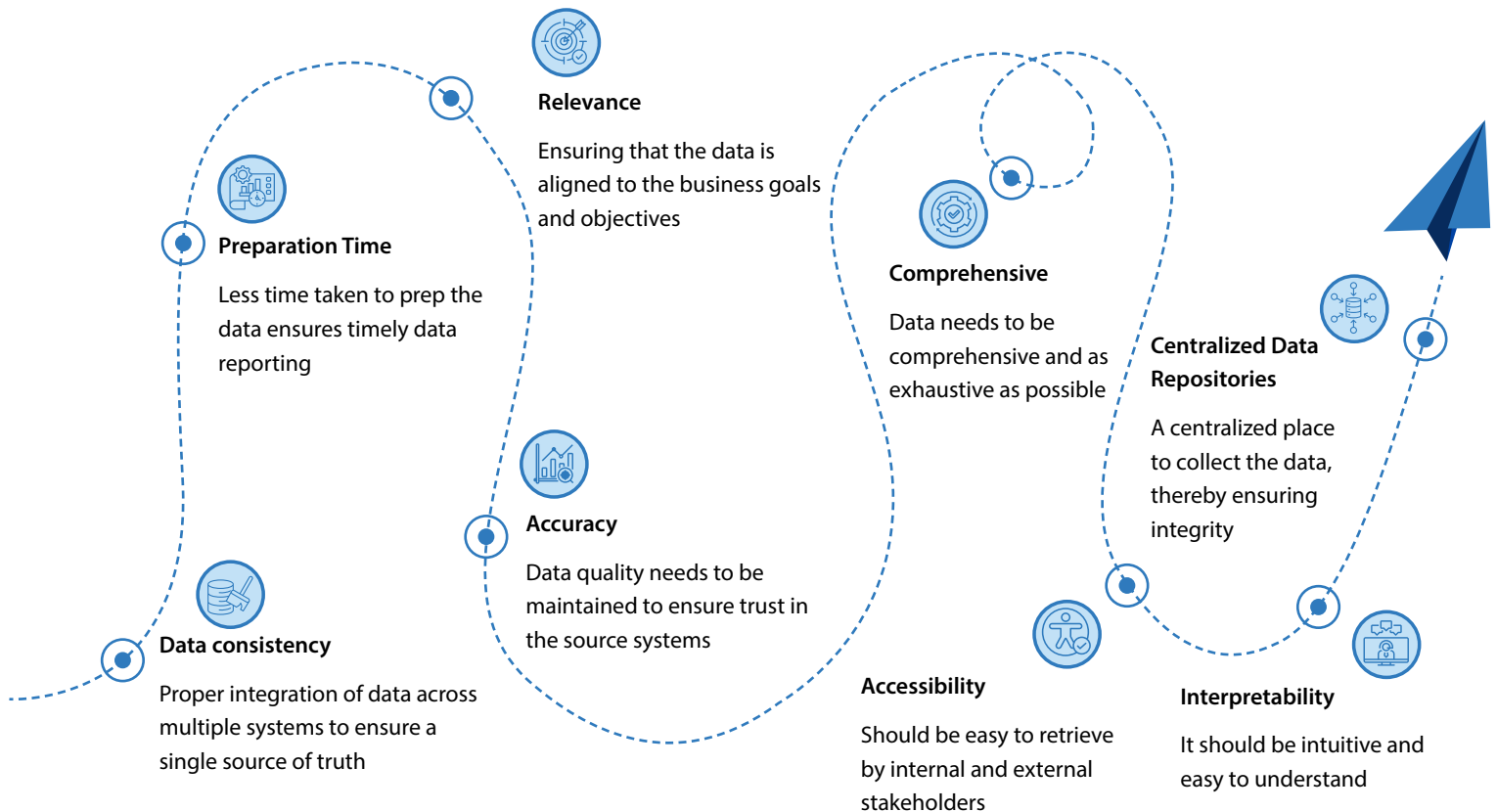Data value increases exponentially through progressive processing and contextual refinement. Raw data transforms into information through systematic contextualization, where data monitoring and reporting mechanisms serve as primary information generation sources. This transformation process is fundamental to enterprise data-driven strategies, as it converts disparate data points into coherent, actionable insights that directly influence strategic business outcomes.

The value of data is fundamental to an enterprise's data-driven approach because it transforms raw information into actionable insights that directly influence business success. The value of data lies in its ability to provide meaningful, trustworthy insights that empower enterprises to make better decisions, optimize operations, innovate, and achieve sustainable growth. Without valuable data, the data-driven approach loses its effectiveness and can lead to misguided or inefficient outcomes.

Some of the factors that contribute to determining the value of data:

Relevance

Ensuring that the data is aligned to the business goals and objectives

Preparation Time

Less time taken to prep the data ensures timely data reporting

Comprehensive

Data needs to be comprehensive and as exhaustive as possible

Centralized Data Repositories

A centralized place to collect the data, thereby ensuring integrity

Accuracy

Data quality needs to be maintained to ensure trust in the source systems

Data consistency

Proper integration of data across multiple systems to ensure a single source of truth

Accessibility

Should be easy to retrieve by internal and external stakeholders

Interpretability

It should be intuitive and easy to understand

The true value of data is realized only when it is properly governed, secured, and integrated across the organization. Enterprise Data Management (EDM) ensures data quality, accessibility, and compliance, making data a reliable asset for decision-making and innovation. Data that is accurate, consistent, and well-managed helps streamline processes, reduce errors, and ensure compliance with regulations, thereby lowering risks and operational costs

Volume

Volume of data is a foundational element for an enterprise data-driven approach because it enables deeper insights, proactive strategies, operational efficiency, and competitive advantage. However, volume alone is insufficient without proper management and quality assurance to transform data into a strategic asset that drives business success.

Excessive irrelevant or poor-quality data can overwhelm decision-makers and lead to "information paralysis." Effective enterprise data management ensures that large volumes of data are accurate,

consistent, and actionable, maximizing their value for decision-making. Managing large volumes of data effectively equips organizations to scale their data-driven initiatives and respond agilely to changing business environments.

Large volumes of data allow organizations to analyse diverse and detailed datasets, which enhances their ability to anticipate market changes and adapt proactively. This comprehensive data analysis offers a competitive edge by uncovering trends and opportunities that smaller datasets might miss. Properly managing large data volumes requires compliance with data retention policies and regulatory requirements.

Key Factors Affecting Data Volume:

Data Sources
Data volume increases proportionally with data sources.

User Base Size
Increased users lead to a proportional rise in interactions, thereby generating more data. E.g. A global app vs. a local one.

Frequency of Data Generation
Real-time systems generate data continuously. Batch systems generate data periodically.

Business Growth
As operations expand, so does the data footprint. E.g. More transactions, more customers, more analytics.

Granularity of Data
Finer granularity (more detailed data) increases volume. E.g. Logging every user click vs. just session summaries.

Regulatory Requirements
Some industries and domains require extensive data logging and archiving. E.g. Financial services, healthcare, aviation.

Retention Policies
Extended user retention contributes to the accumulation of immaterial data over time. E.g. Keeping activity logs for 1 year vs 1 month.

Data Duplication and Redundancy
Poor data management can lead to unnecessary volume. E.g. Multiple copies of the same dataset across systems.

Data Types and Formats
Rich formats (images, video, audio) consume more space than text or numeric data. E.g. Surveillance footage vs. temperature readings.

Machine Learning and Analytics
Training models and storing intermediate results can significantly increase volume. E.g. Feature stores, model logs, experiment tracking.

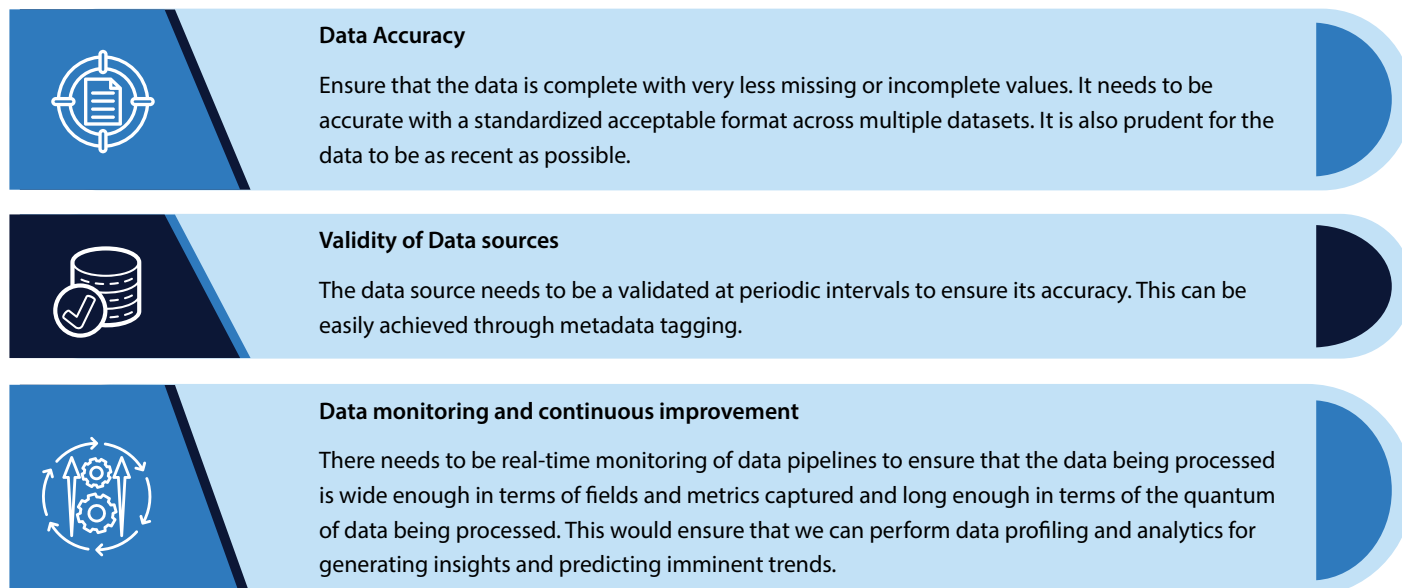# Key business areas/levers where IT organizations can leverage Data

Key archetypes/levers across which a Data-driven IT Approach could be leveraged:

1) Ensuring Data Trustworthiness

Improving data trustworthiness is essential for making key business decisions that would drive the goals and outcomes of an IT organization. Some of the key parameters to consider are:

Data Accuracy

Ensure that the data is complete with very less missing or incomplete values. It needs to be accurate with a standardized acceptable format across multiple datasets. It is also prudent for the data to be as recent as possible.

Validity of Data sources

The data source needs to be a validated at periodic intervals to ensure its accuracy. This can be easily achieved through metadata tagging.

Data monitoring and continuous improvement

There needs to be real-time monitoring of data pipelines to ensure that the data being processed is wide enough in terms of fields and metrics captured and long enough in terms of the quantum of data being processed. This would ensure that we can perform data profiling and analytics for generating insights and predicting imminent trends.

2) Improving Data Quality and governance through a designated data operating model

Having a clear Data Strategy and Vision helps you to business goals and the broader organizational mission. Some of the key facets are:

Data Governance

establishes rules, policies, and standards to ensure data quality, security, and compliance and it defines roles like data stewards and ensures accountability for data handling.

Data Quality

This is the assessment of data's fitness to serve its purpose. It is determined by factors such as accuracy, completeness, reliability, relevance and timeliness.

Data Architecture

This involves the organization of data elements that drives efficient and secure data transmission between systems.

Data Management

Managing meta-data comprising of business, technical and operational information. This provides clarity in data relationship and data lineage

Data Security

This refers to the compliance and privacy needs of data throughout its lifecycle. It includes data identity, access management, encryption and the best practices associated with it.

Data Operations

This involves technical and functional operations that utilizes the use of data platforms after go-live as well as involvement of future business owners.

3) Fostering data driven organization culture and decision making

CEO's and CIOs of the world to make decisions that are guided by evidence and insights derived from data rather than intuition or assumptions. There is a clear sense of accountability wherein clear metrics and KPIs are established, helping teams track performance and take responsibility for outcomes. Some of these measures are:

Continuous Improvement

A mindset of using data to refine processes, optimize strategies, and adapt to change is fostered throughout. All in all, it fosters a data-driven culture where employees understand the value of data. This also encourages the

Adoption of data-centric practices

It creates transparency wherein data is shared across teams to create an environment of openness and collaboration, eliminating silos. Data needs to be the cornerstone of any planning and budgeting cycle in an IT organization.

Holistic data architecture

This provides a blueprint for how data is structured, integrated, and accessed within the organization. Using the right historical data captured from the financial General Ledger and implementing forecasting models as Time series – ARIMA (e.g.), one can successfully plan for the resources and downstream revenue pipeline. Customer lead generation and key account growth are some of the benefits. This ensures alignment with business needs and scalability for future growth.

4) Amplifying Data Security

Enterprises should not just ensure adherence to regulations, to avoid fines and mitigate legal risk but should also aim to build stakeholder trust by adopting Security-First culture. Some of the best practices are:

Applying Security Controls

With streamlined operations, well-defined data security and privacy controls like data anonymization and unified data security and governance framework, enterprises can fuel their growth. Regulatory standards such as GDPR, CCPA, HIPAA, and PCI-DSS mandate robust data protection measures. Adhering to these regulations is crucial not only to avoid legal penalties but also to demonstrate organizational accountability and commitment to data privacy.
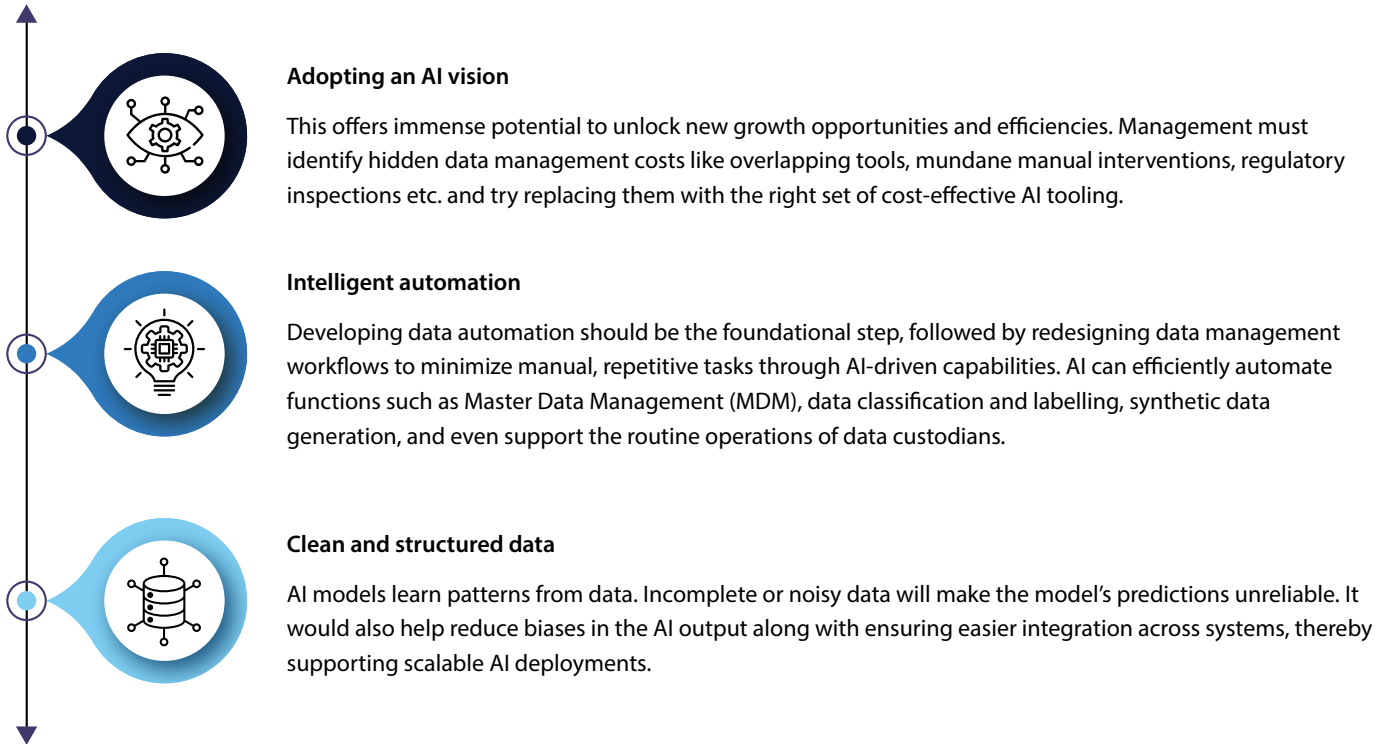
Ensuring Operational Continuity

Modern cybersecurity technologies such as Zero Trust Network Access (ZTNA), Cloud Access Security Brokers (CASB), Next-Generation Firewalls (NGFW) etc. are critical for defence against evolving cyberattack methodologies. Right mix of proactive organizational changes and technical solutioning is essential across cloud, on-premises, and hybrid environments for rapidly expanding data landscapes.

The future of financial crime detection unequivocally lies in embracing these advanced technologies. By adopting and constantly evolving their Graph AI solutions, financial institutions can not only stay significantly ahead of evolving threats but also ensure robust security, maintain stringent compliance, and foster greater trust in an increasingly interconnected and complex global financial landscape. This strategic foresight will be a key differentiator in the competitive financial services market in the months and years ahead.
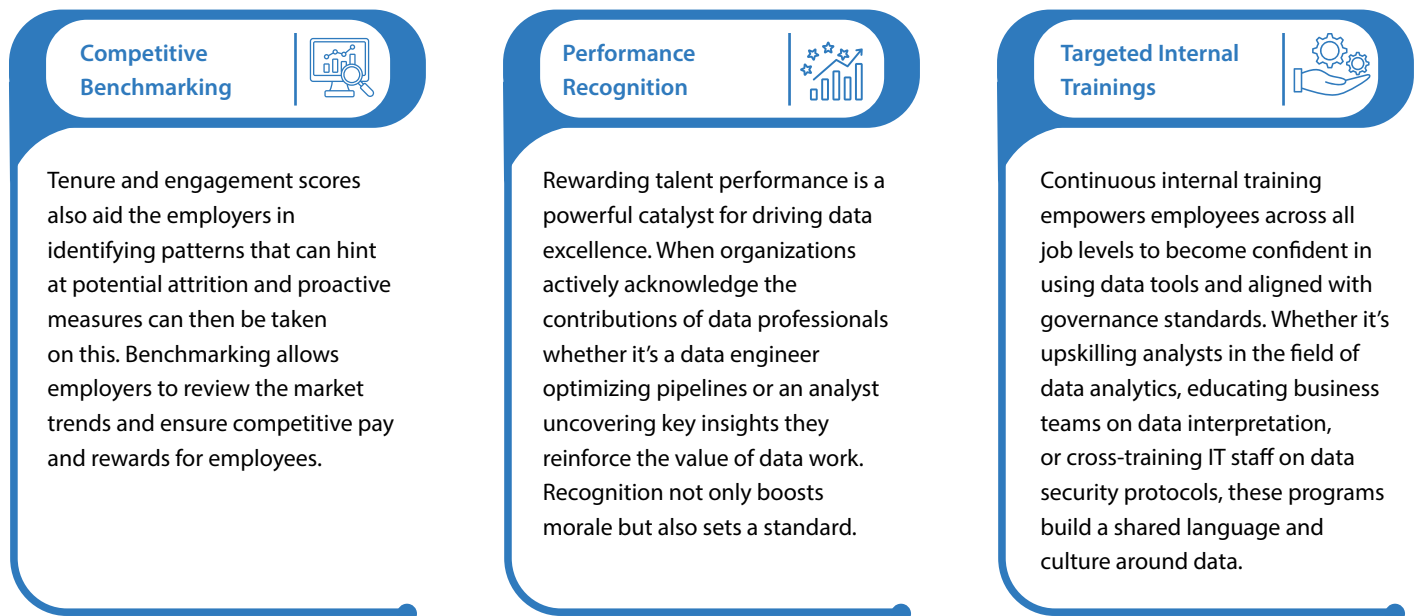
5) Creation of a Data Ecosystem with AI-enabled functions

With the use of GenAI and Agentic AI, enterprises can simplify and supplement their data management tasks to accelerate time to value.

Adopting an AI vision

This offers immense potential to unlock new growth opportunities and efficiencies. Management must identify hidden data management costs like overlapping tools, mundane manual interventions, regulatory inspections etc. and try replacing them with the right set of cost-effective AI tooling.

Intelligent automation

Developing data automation should be the foundational step, followed by redesigning data management workflows to minimize manual, repetitive tasks through AI-driven capabilities. AI can efficiently automate functions such as Master Data Management (MDM), data classification and labelling, synthetic data generation, and even support the routine operations of data custodians.

Clean and structured data

AI models learn patterns from data. Incomplete or noisy data will make the model's predictions unreliable. It would also help reduce biases in the AI output along with ensuring easier integration across systems, thereby supporting scalable AI deployments.

6) Talent Retention and Employee Engagement

Data helps you to enhance employee engagement by capturing data from performance reviews and surveys that helps you identify the most valued facets by the employees. This can be realised through following initiatives:

Competitive Benchmarking

Tenure and engagement scores also aid the employers in identifying patterns that can hint at potential attrition and proactive measures can then be taken on this. Benchmarking allows employers to review the market trends and ensure competitive pay and rewards for employees.

Performance Recognition

Rewarding talent performance is a powerful catalyst for driving data excellence. When organizations actively acknowledge the contributions of data professionals whether it's a data engineer optimizing pipelines or an analyst uncovering key insights they reinforce the value of data work. Recognition not only boosts morale but also sets a standard.

Targeted Internal Trainings

Continuous internal training empowers employees across all job levels to become confident in using data tools and aligned with governance standards. Whether it's upskilling analysts in the field of data analytics, educating business teams on data interpretation, or cross-training IT staff on data security protocols, these programs build a shared language and culture around data.

With the right mix of absorbing the right outside-in talent from the industry as well as creating a culture of continuous improvement through upskilling helps in establishing a data-driven organization.

Conclusion

Data transcends its traditional role as an operational resource to become the fundamental enabler of AI transformation. It serves as the essential substrate through which machines acquire learning capabilities, adapt to evolving conditions, and execute intelligent decision-making processes. AI-driven models are inherently dependent on high-quality data to recognize complex patterns, optimize algorithmic performance, and continuously enhance predictive accuracy. Without robust data foundations, AI systems cannot achieve their intended functionality, evolutionary capacity, or strategic value delivery.

Organizations that strategically position data as a core business asset will establish competitive advantages in responsible AI innovation. By systematically addressing data management through comprehensive frameworks—encompassing volume, velocity, variety, veracity, and value considerations organizations can maintain operational agility, market competitiveness, and strategic preparedness for future technological developments.

The cultivation of an organizational culture that prioritizes data quality, governance excellence, and innovation-driven practices, combined with the strategic integration of AI capabilities into data management workflows, enables organizations to harness AI's transformative potential. This approach facilitates enhanced decision-making processes, delivers personalized customer experiences, and drives operational excellence in the digital economy. Through this comprehensive data-centric strategy, Chief Data Officers can establish future-resilient enterprises while serving as catalysts for sustained innovation, operational resilience, and strategic growth in an increasingly data-driven business landscape.

About the Authors

Raj Nellutla

Raj is a highly accomplished Data Strategy & Digital Transformation Executive Leader who implements new business strategies for global companies. He is well known as Strategic thinker and Adviser to CIO and CDO, adept in aligning with the corporate leadership goals while developing and launching new strategies.

Tushar Arya

Tushar Arya is a Principal with the Tech Transformation practice of Infosys Consulting. He carries extensive experience in leading IT and Data Strategy engagements for multiple organizations across domains of BFSI, Oil & Gas, Telecom and Utilities.

Ayush Dabas

Ayush Dabas is a Senior Consultant with the Tech Transformation practice of Infosys Consulting. He has rich consulting experience of more than 6 years in Cybersecurity Strategy, IT Audits and Data Security.

For more information, contact askus@infosys.com

Infosys®
Navigate your next

Infosys.com | NYSE: INFY

Stay Connected