Journal of Emerging Intelligence and Engineering Technologies

Research Article

Integrating AI, Big Data, and Data Intelligence for Next-Generation Enterprise Systems

Rajan Bhattacharya

Universiti Kebangsaan Malaysia

Abstract

Artificial intelligence (AI), big data, and data intelligence have each matured substantially as independent disciplines, yet their greatest organizational value emerges when they are architecturally integrated rather than deployed in isolation: big data infrastructure supplies the scale and variety of information required to train and ground AI models, AI provides the analytical and generative capability to extract patterns and generate recommendations from that data, and data intelligence supplies the governance, quality, and organizational practices required to ensure resulting insight is trustworthy and actionable. This paper presents a comprehensive review of the integration of AI, big data, and data intelligence into next-generation enterprise systems, examining the converged reference architecture spanning data infrastructure, AI/ML processing, and decision intelligence layers, and quantifying current enterprise adoption patterns and market trajectories using recent industry survey and market-forecast data. The review surveys representative integration patterns, including AI-augmented data pipelines, intelligent data fabrics, and closed-loop enterprise decision systems, and examines applications across manufacturing, financial services, healthcare, and retail. Critical open challenges are discussed, including architectural and organizational silos between data and AI teams, data quality as a bottleneck for AI reliability, integration cost and legacy system compatibility, governance of increasingly autonomous AI-driven data systems, and workforce and change-management gaps. The paper concludes by identifying emerging directions, including unified data-AI platforms, agentic enterprise data intelligence, and real-time intelligent data fabrics. This review is intended to serve as a consolidated, evidence-based reference for researchers and practitioners designing integrated, next-generation enterprise systems.

Keywords: Artificial Intelligence, Big Data, Data Intelligence, Enterprise Systems, Data Fabric, AI Integration, Digital Transformation, Enterprise Architecture

Received: 10 July 2025
Revised: 03 Au1ust 2025
Accepted: 15 November 2025
Published: 30 December2025

Introduction

Artificial intelligence (AI), big data, and data intelligence have each developed into mature, extensively researched disciplines in their own right. Big data technologies provide the infrastructure to capture, store, and process information at a scale and variety unattainable by traditional database systems. Artificial intelligence, particularly machine learning and, more recently, generative AI, provides the analytical and reasoning capability to extract patterns, generate predictions, and increasingly produce natural-language insight from that data. Data intelligence, as reviewed in related literature, supplies the governance, quality assurance, and organizational practices required to ensure that data and AI-derived output are trustworthy, compliant, and genuinely actionable in decision-making.

While each discipline has been extensively studied in isolation, a growing body

of research and industry practice emphasizes that the greatest enterprise value emerges not from any single discipline but from their deliberate architectural and organizational integration. An enterprise with extensive big data infrastructure but weak AI capability cannot translate its data volume into predictive or generative insight; an enterprise with sophisticated AI models but poor underlying data quality and governance will produce unreliable or untrustworthy output regardless of model sophistication; and an enterprise lacking a coherent data intelligence layer will struggle to translate either data or AI capability into consistent, trusted organizational decision-making. This paper refers to the deliberate architectural integration of these three disciplines as the foundation of next-generation enterprise systems.

Recent industry survey data underscores both the scale of this integration trend and its uneven realization across organizations. According to McKinsey's 2025-26 State of AI survey, 44 percent of surveyed organizations now report scaling AI across the enterprise, up from 38 percent the prior year, though this capability remains substantially concentrated among larger organizations: 54 percent of enterprises with revenue exceeding one billion US dollars report enterprise-wide AI scaling, compared with only 33 percent of smaller organizations, a pattern that recurs for AI agent scaling specifically (40 percent versus 22 percent). This scale disparity suggests that the technical and organizational demands of integrating AI, big data, and data intelligence remain a differentiating capability rather than a uniformly distributed one.

This paper presents a comprehensive review of the integration of AI, big data, and data intelligence for next-generation enterprise systems, with the following objectives: (1) to present a converged reference architecture spanning the three disciplines; (2) to quantify current enterprise adoption patterns and market trajectories using recent survey and market-forecast data; (3) to survey representative integration patterns and application domains; (4) to critically examine open challenges; and (5) to identify promising future research and practice directions.

The remainder of this paper is organized as follows. Section 2 presents the converged reference architecture. Section 3 examines current adoption and market data. Section 4 surveys integration patterns and application domains. Section 5 discusses open challenges. Section 6 outlines future directions, and Section 7 concludes the paper.

2. A Converged Reference Architecture

Next-generation enterprise systems can be understood as a layered architecture in which big data, AI, and data intelligence each occupy a distinct but interdependent layer, connected by a cross-cutting integration fabric. The data infrastructure layer, grounded in big data technologies, ingests, stores, and processes structured, semi-structured, and unstructured data at scale, drawing on data lakehouse architectures, distributed and streaming processing frameworks, and increasingly cloud-native, elastically scalable storage and compute. This layer's primary responsibility is ensuring that data of sufficient volume, variety, and freshness is available to support both traditional analytics and AI model training and inference.

The AI/ML processing layer consumes data from the infrastructure layer to train, fine-tune, and serve machine learning and generative AI models, ranging from classical predictive models to large foundation models accessed via API or deployed on-premise, orchestrated through MLOps pipelines that manage model versioning, deployment, and monitoring. The decision intelligence layer, corresponding to the data intelligence discipline, governs and contextualizes the output of the AI layer, applying data quality assurance, semantic and business-context enrichment, and governance controls before insight is delivered to decision-makers or triggers automated action.

Table 1: Converged Reference Architecture for AI, Big Data, and Data Intelligence Integration

Architecture Layer	*Primary Discipline*	*Core Function*	*Representative Technologies*
Data Infrastructure Layer	Big Data	Ingests, stores, and processes structured/unstructured data at scale	Data lakehouses, distributed processing, streaming platforms
AI/ML Processing Layer	Artificial Intelligence	Extracts patterns, predictions, and generative output from data	ML/DL models, foundation models, MLOps pipelines
Decision Intelligence Layer	Data Intelligence	Governs, contextualizes, and operationalizes insight into decisions	Governance frameworks, semantic layers, decision engines
Integration Fabric	Cross-Cutting	Connects layers via metadata, APIs, and unified access	Data fabric, knowledge graphs, enterprise APIs

Binding these three layers together, the integration fabric, increasingly conceptualized in industry practice as a data fabric or data mesh architecture, provides unified metadata management, semantic consistency, and API-level connectivity across otherwise heterogeneous underlying systems, enabling data, models, and insight to flow coherently across the architecture rather than remaining siloed within each discipline's traditional tooling. Knowledge graphs increasingly serve a related integrative function, providing a shared semantic layer that both AI models and human decision-makers can query consistently.

3. Enterprise Adoption and Market Trajectory

Quantitative evidence from recent industry surveys and market analyses illustrates both the pace of integration and the substantial scale disparity in its realization. Figure 1 presents enterprise AI scaling data from McKinsey's 2025-26 State of AI survey, disaggregated by organization size. Large enterprises with revenue exceeding one billion US dollars report substantially higher rates of both enterprise-wide AI scaling (54 percent) and AI agent scaling (40 percent) relative to smaller organizations (33 percent and 22 percent respectively), a gap

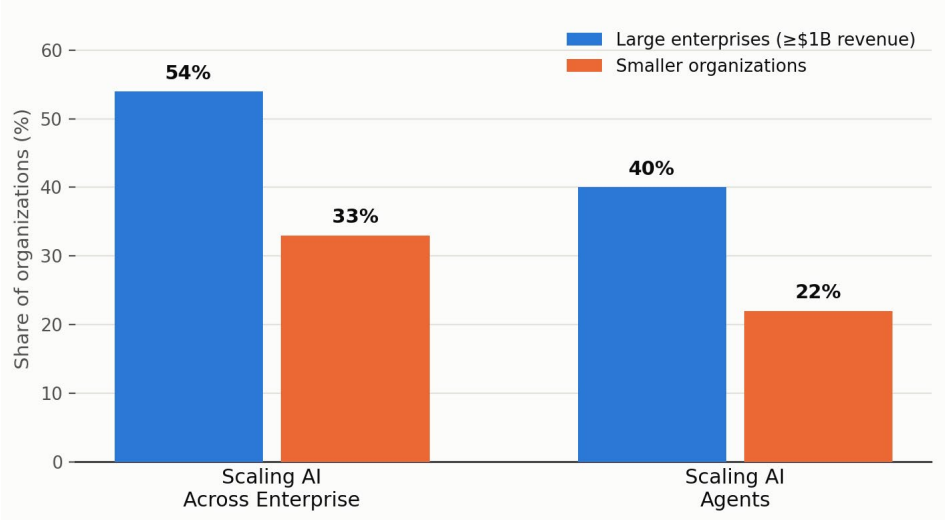

Figure 1: Enterprise AI scaling by organization size, 2025–26. Source: McKinsey State of AI Survey.

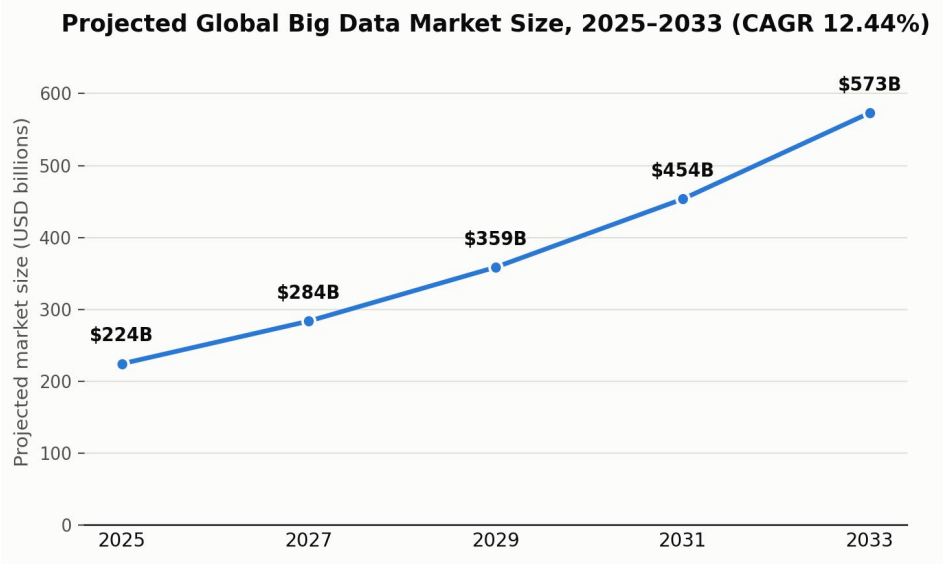

Figure 2: Projected global big data market size, 2025–2033. Source: market forecast data compiled from industry analyses.

that industry analysts consistently attribute to the greater data infrastructure maturity, specialized talent availability, and capital investment capacity of larger organizations, each a prerequisite for effective AI-big data-data intelligence integration at scale.

The underlying data infrastructure that supports this AI integration continues to expand rapidly. Figure 2 presents the projected trajectory of the global big data market from 2025 through 2033, based on a reported 2025 market size of approximately $224.46 billion growing at a compound annual growth rate (CAGR) of 12.44 percent to a projected $573.47 billion by 2033, reflecting sustained enterprise investment in the data infrastructure layer that underpins AI and data intelligence capability. Adjacent market segments show comparable or steeper growth trajectories, with the broader big data analytics market projected to grow from $394.70 billion in 2025 to $1,176.57 billion by 2034 (CAGR 12.80 percent), and the big data-as-a-service segment specifically projected to grow at a CAGR of 19.11 percent, reflecting particularly strong enterprise demand for managed, cloud-delivered data infrastructure that lowers the technical barrier to integration for organizations lacking in-house data engineering capacity.

Table 2 summarizes market size and growth projections across several related big data and analytics market segments, illustrating the breadth of infrastructure investment underlying enterprise AI and data intelligence integration.

Table 2: Big Data and Analytics Market Size and Growth Projections

Market Segment	*2025 Market Size (USD)*	*Forecast Market Size (USD)*	*CAGR*
Global Big Data Market	$224.46 billion (2025)	$573.47 billion (2033)	12.44%
Big Data Analytics Market	$394.70 billion (2025)	$1,176.57 billion (2034)	12.80%
Big Data & Analytics Services	—	$440.03 billion (2030)	21.50%
Big Data Platform Market	— (2026: $101.55B)	$314.35 billion (2035)	13.38%
Big Data as a Service (BDaaS)	— (2026: $47.21B)	$226.81 billion (2035)	19.11%

Integration Patterns and Application Domains

Integration Patterns

Several recurring patterns characterize how organizations architecturally integrate AI, big data, and data intelligence in practice. AI-augmented data pipelines embed machine learning directly within data ingestion and transformation workflows, for example using AI-based data quality anomaly detection or automated schema matching, blurring the traditional boundary between data engineering and AI functions. Intelligent data fabrics extend traditional data fabric architectures with embedded AI capability, enabling automated data discovery, classification, and semantic enrichment across an organization's full data estate, substantially reducing the manual data preparation effort historically required before AI models can be trained or deployed against enterprise data. Closed-loop enterprise decision systems integrate all three layers end-to-end, from data ingestion through AI-driven prediction or generation to governed, automated or human-reviewed decision execution, representing the most architecturally mature and organizationally impactful integration pattern, though one that remains concentrated among the larger, more data-mature organizations reflected in Figure 1.

Representative Application Domains

Manufacturing: Integrated systems combine IoT sensor data infrastructure, AI-driven predictive maintenance and quality inspection models, and data intelligence-governed maintenance scheduling and resource allocation decisions, increasingly coordinated with digital twin representations of physical assets and processes.

Financial Services: Financial institutions integrate high-velocity transactional data infrastructure with AI-driven fraud detection, credit risk, and algorithmic trading models, governed by data intelligence frameworks addressing the stringent regulatory, auditability, and explainability requirements characteristic of the sector.

Healthcare: Integrated systems combine electronic health record and medical imaging data infrastructure with AI-driven diagnostic and risk-prediction models, governed by data intelligence frameworks that must additionally satisfy healthcare-specific privacy regulation and clinical validation requirements before insight informs patient care decisions.

Retail: Retailers integrate omnichannel transactional and customer behavior data infrastructure with AI-driven demand forecasting, personalization, and pricing models, governed by data intelligence practices increasingly extended to address AI-specific concerns such as recommendation fairness and customer data privacy across integrated online and physical retail channels.

Open Challenges

Organizational and Architectural Silos: Data engineering, AI/ML, and data governance functions have historically developed as organizationally and technically distinct disciplines with separate tooling, teams, and reporting lines, and dismantling these silos to enable the architectural integration described in Section 2 remains a substantial organizational change management challenge that frequently outpaces the underlying technical integration effort.

Data Quality as an AI Reliability Bottleneck: As AI models become more deeply embedded within data pipelines and enterprise decision systems, the reliability of AI output becomes increasingly bounded by the quality of the underlying big data infrastructure, reinforcing rather than reducing the criticality of data quality and governance practices even as AI capability advances.

Integration Cost and Legacy System Compatibility: The market data presented

in Section 3 reflects substantial ongoing infrastructure investment, but industry reporting consistently identifies integration with legacy enterprise systems, rather than the acquisition of new data or AI technology per se, as a leading cost and complexity driver in realizing integrated next-generation enterprise architectures.
Governance of Increasingly Autonomous Systems: As closed-loop decision systems integrating all three layers become more prevalent, particularly those incorporating increasingly autonomous AI agents, governance frameworks must evolve to address not only data quality and AI model risk in isolation but the emergent risk of increasingly autonomous, integrated systems acting on enterprise data with limited human review, an area in which governance practice reportedly lags technical capability.
Scale Disparity and the Widening Adoption Gap: The substantial gap in AI and data integration maturity between large and smaller organizations documented in Figure 1 raises concerns regarding a widening competitive disparity, as smaller organizations face proportionally greater barriers to the capital investment, specialized talent, and data infrastructure maturity required for full integration.
Workforce and Change Management: Consistent with broader data intelligence and enterprise AI literature, realizing value from integrated AI-big data-data intelligence systems depends substantially on workforce data and AI literacy and organizational change management, dimensions that technical architecture alone cannot resolve.

Future Directions

Unified Data-AI Platforms: Continued convergence of previously separate data infrastructure and AI/ML platform tooling into unified enterprise platforms is likely to reduce the integration friction and organizational silos identified in Section 5, lowering the technical barrier to end-to-end architectural integration.
Agentic Enterprise Data Intelligence: The extension of agentic AI architectures across the full integrated stack, enabling AI agents to autonomously query, analyze, and act on enterprise data within governed bounds, represents a rapidly emerging direction with significant implications for both the productivity potential and governance complexity of next-generation enterprise systems.
Real-Time Intelligent Data Fabrics: Continued advancement of low-latency, AI-augmented data fabric architectures is likely to expand the range of enterprise use cases for which integrated, real-time predictive and prescriptive decision-making is feasible, extending current capability beyond the batch-oriented integration patterns that remain common today.
Democratization of Integration Capability: Continued growth of managed, cloud-delivered big data and AI platform services, reflected in the particularly strong projected growth of the big data-as-a-service market segment, is likely to gradually narrow the scale disparity documented in Figure 1 by lowering the infrastructure and talent barriers to integration for smaller organizations.
Standardized Governance Frameworks for Integrated Systems: The development of governance frameworks explicitly addressing the emergent risks of integrated, increasingly autonomous data-AI systems, rather than treating data governance and AI governance as separate concerns, represents an important direction for enabling responsible adoption of closed-loop enterprise decision systems at scale.

Conclusion

The integration of artificial intelligence, big data, and data intelligence represents a defining architectural and organizational challenge for next-generation enterprise systems, extending the value of each individual discipline substantially beyond what it can achieve in isolation. This review has presented a converged reference architecture spanning data infrastructure, AI/ML processing, and decision

intelligence layers, connected through an integration fabric, and has grounded this architectural discussion in current quantitative evidence: enterprise AI scaling remains substantially concentrated among larger organizations, even as the underlying big data market continues to grow at a compound annual rate exceeding twelve percent toward a projected $573 billion by 2033. While representative integration patterns and applications across manufacturing, financial services, healthcare, and retail demonstrate the feasibility and value of integrated architectures, persistent challenges around organizational silos, data quality, legacy integration cost, governance of autonomous systems, and the widening scale disparity between large and smaller organizations remain significant barriers to broader realization. Continued progress toward unified data-AI platforms, agentic enterprise data intelligence, and democratized, cloud-delivered integration capability is likely to define the next phase of enterprise system architecture.

References

[1] H. Chen, R. H. L. Chiang, and V. C. Storey, "Business Intelligence and Analytics: From Big Data to Big Impact," MIS Quarterly, vol. 36, no. 4, pp. 1165-1188, 2012.

[2] A. Gandomi and M. Haider, "Beyond the hype: Big data concepts, methods, and analytics," International Journal of Information Management, vol. 35, no. 2, pp. 137-144, 2015.

[3] D. Sculley, G. Holt, D. Golovin, E. Davydov, T. Phillips, D. Ebner, et al., "Hidden Technical Debt in Machine Learning Systems," in Proc. Advances in Neural Information Processing Systems (NeurIPS), 2015, pp. 2503-2511.

[4] McKinsey & Company, "The State of AI: Global Survey 2025-26," McKinsey QuantumBlack, 2026. Available: https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai

[5] Itransition, "Future of Big Data: Forecasts, Statistics and Trends for 2026," Itransition Research, 2025. Available: https://www.itransition.com/data/big/future

[6] Deloitte, "The State of AI in the Enterprise," Deloitte Insights, 2026. Available: https://www.deloitte.com/us/en/what-we-do/capabilities/applied-artificial-intelligence/content/state-of-ai-in-the-enterprise.html

[7] Indata Labs, "Technology Trends 2025: AI and Big Data Analytics," Indata Labs Report, 2025. Available: https://indatalabs.com/wp-content/uploads/reports/technology-trends-2025-ai-and-big-data-analytics.pdf

[8] KPMG International, "Data Governance in the Age of AI," KPMG Insights, 2025. Available: https://kpmg.com/kpmg-us/content/dam/kpmg/pdf/2025/data-governance-age-ai.pdf

[9] OpenAI, "The State of Enterprise AI, 2025 Report," 2025. Available: https://cdn.openai.com/pdf/7ef17d82-96bf-4dd1-9df2-228f7f377a29/the-state-of-enterprise-ai_2025-report.pdf

[10] Menlo Ventures, "2025: The State of Generative AI in the Enterprise," Menlo Ventures Perspectives, 2025. Available: https://menlovc.com/perspective/2025-the-state-of-generative-ai-in-the-enterprise/