June 2026
White Paper
Sponsored by Everpure

Why AI Projects Fail: Revealing the Data Infrastructure Gap

Phil Goodwin
Research Vice President,
Infrastructure Software Platforms,
Worldwide Infrastructure Research, IDC

Table of contents

Click any title to navigate directly to that page.

IDC opinion

Artificial intelligence has become the number 1 investment priority for organizations worldwide. Enterprises are aggressively pursuing AI initiatives because of the substantial potential benefits: improved customer experience, more efficient staff, accelerated business processes, and, ultimately, stronger financial performance.

Yet, for all the promise and momentum surrounding AI, these projects too often fail to deliver tangible business value. IDC research reveals that more than half (54%) of AI proof-of-concept projects fail to advance to production. This means that the majority of AI project budgets generate zero return on investment.

As organizations grow more experienced and budget conscious, they are moving away from trial-and-error approaches to AI investment. They are becoming more selective about which projects receive funding and more deliberate about what conditions must be met for those projects to succeed.

54%
of AI proof-of-concept projects fail to advance to production, generating zero ROI for the organization.

IT teams are learning that data-related issues are chief among the factors that determine AI project outcomes. These critical success factors include data availability without silos, data timeliness and currency, and robust data governance. Simply stated, no AI project can succeed without the right data.

IDC research finding

94% of survey respondents state that data quality is important or very important to AI project success. The top 3 contributors to poor data quality are:

- Redundant data
- Multiple storage systems (silos)
- Obsolete data

This reality does not diminish the importance of storage systems capable of optimizing GPU clusters and addressing the unique performance demands of AI workloads. IDC research confirms that legacy storage infrastructure is failing to meet those demands, specifically in low-latency access and high-throughput data delivery.

Because of the shortcomings endemic in current storage architectures, many IT organizations are actively re-evaluating their systems against the requirements of AI workloads. IDC refers to the class of systems designed to meet these requirements as AI-ready data storage infrastructure (AI-RDSI).

Modernizing to AI-RDSI is not a onetime purchase or event. It is an ongoing maturation process one that involves people, processes, and technology evolving together. Like most maturation processes, it begins with emergent, largely manual efforts and progresses toward highly automated, repeatable, and future-flexible architectures.

This white paper describes the essential elements of AI-ready data storage infrastructure, presents IDC's AI Readiness Index maturity framework, and outlines the steps organizations can take to advance to the highest tier of readiness.

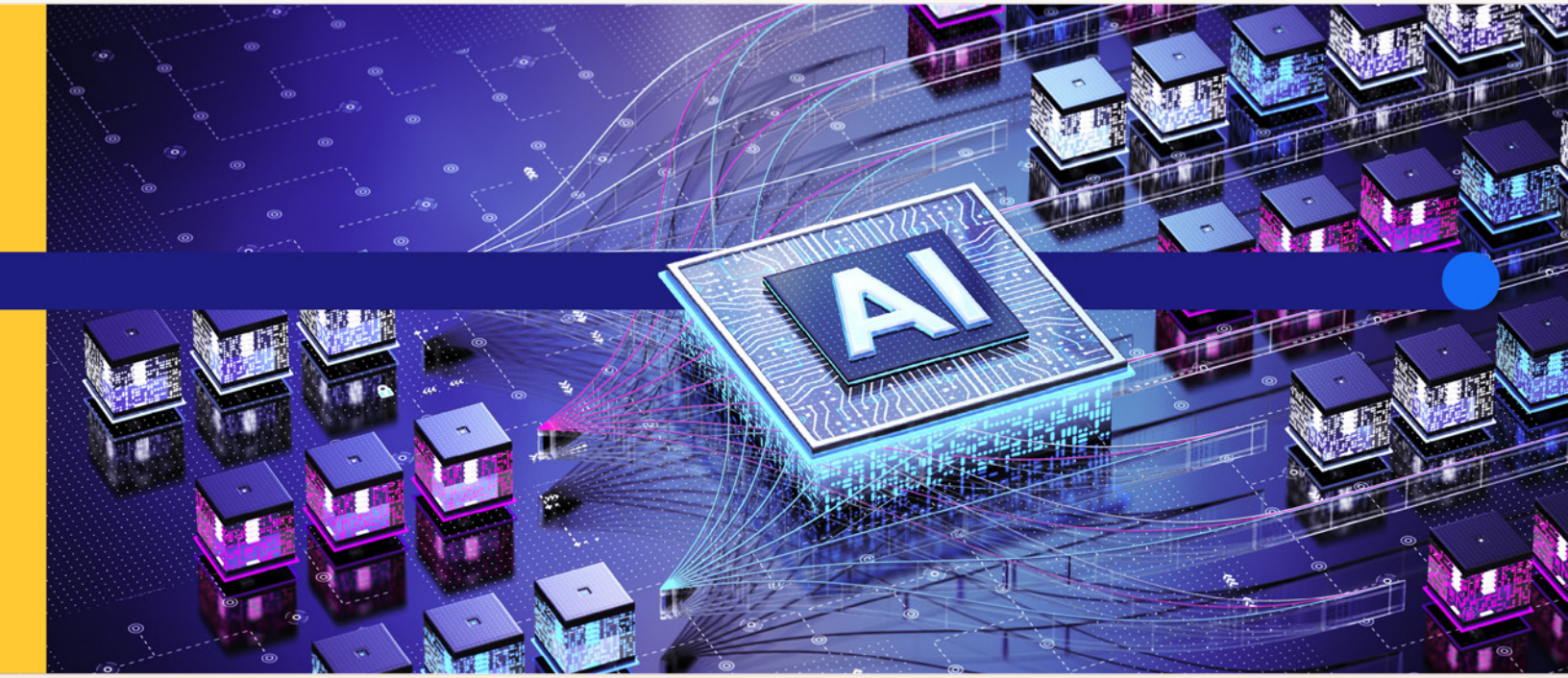
Methodology

IDC conducted a global survey, sponsored by Everpure, of more than 1,300 organizations to develop the findings presented in this paper. The survey included representative samples across countries, industries, business sizes, and respondent roles, ensuring that the results reflect a broad and balanced cross-section of organizational experience.

The survey focused specifically on IT activities related to AI workloads, with particular emphasis on data storage, and on the business outcomes associated with those activities. Using statistical analysis, IDC correlated specific IT practices with measurable business results.

By stratifying these results across performance tiers, IDC identified four distinct groups of organizations, ranging from those in early, exploratory stages (Experimenting) to those operating with fully optimized, future-ready AI data infrastructure (Masters). This framework constitutes the IDC AI Readiness Index.

The AI Readiness Index, along with an associated AI-readiness assessment tool, provides IT leaders with a structured way to identify where their organization stands on the AI data infrastructure maturity journey and the concrete milestones required to advance to the next stage.



Situation overview

Defining AI-ready data storage infrastructure

The hardware, software, and services necessary to prepare, ingest, store, classify, manage, protect, secure, govern, and move data to address the requirements of artificial intelligence applications. AI-RDSI also encompasses service levels pertaining to AI workloads, including performance and system availability; data quality–related attributes, such as trust and provenance; and technologies for the disposition of post-analysis data.

This definition reflects the comprehensive scope of what is required to support modern AI workloads, which encompasses not just the physical storage layer but also the full data life cycle, from ingestion through governance, security, and eventual disposition.

AI workload categories

AI workloads vary significantly in their data requirements and infrastructure demands. Four primary AI workload categories must be considered in any infrastructure planning effort:

Generative AI (GenAI):
Large-scale language and content generation models that require extensive unstructured data, high throughput, and substantial storage capacity

Predictive AI:
Models that forecast outcomes based on historical structured and semistructured data, demanding reliable, well-governed data pipelines

Interpretive AI:
Systems that analyze and derive meaning from complex datasets, including image, video, and sensor data, requiring low-latency access to diverse data types

Agentic AI:
Autonomous AI agents that act on behalf of users or orchestrate multi-step processes, placing especially high demands on real-time data access, security, and governance

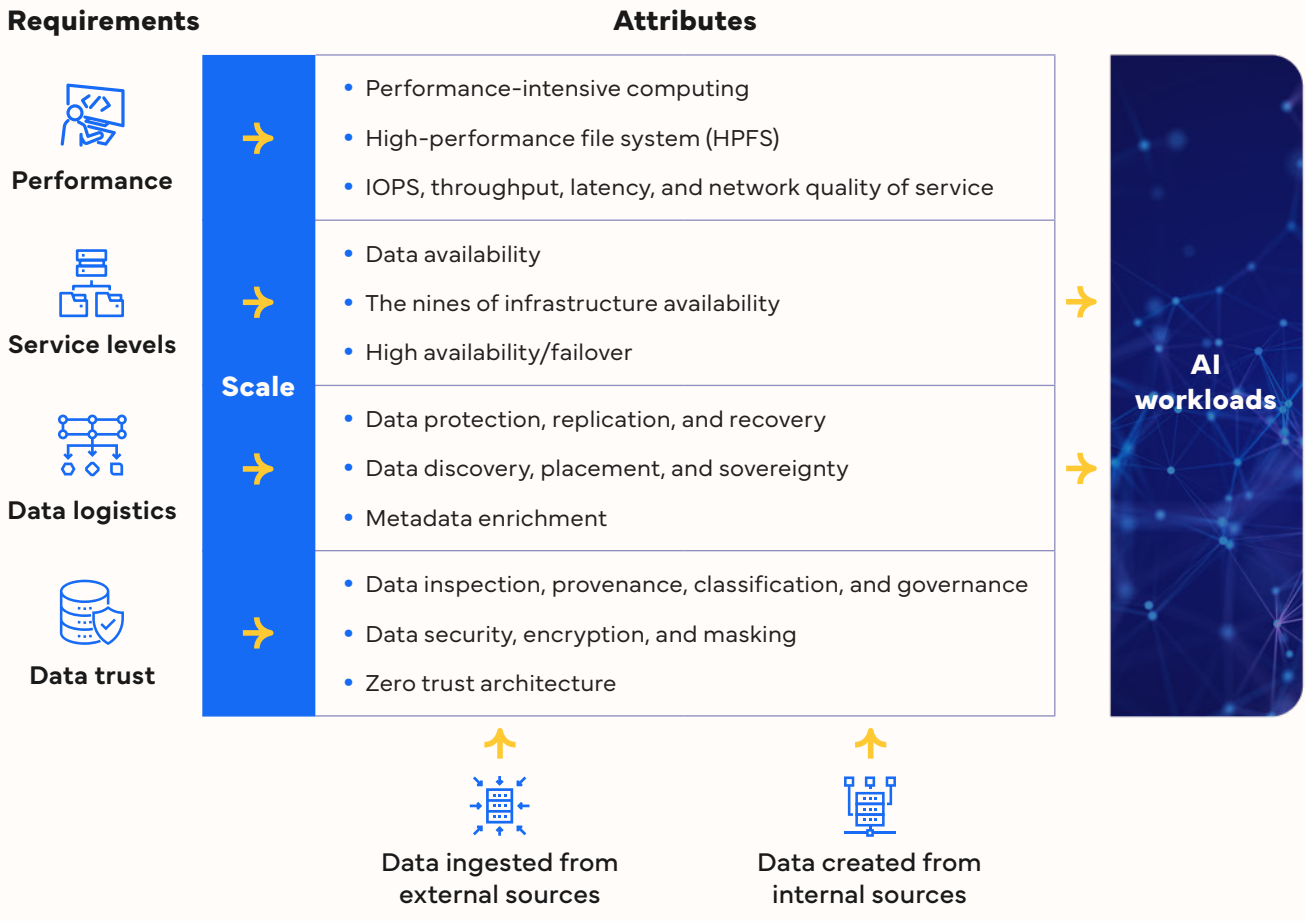
Organizations typically deploy AI using all of the above, with development and deployment occurring across a mix of public cloud and private infrastructure, with no single model or platform dominant. This hybrid reality increases the complexity of data management and amplifies the need for intelligent, context-aware storage systems. Context awareness is enabled by knowledge graphs and semantic layers. These knowledge graphs and semantics allow the systems to adapt to the AI workload requirements and reduce data silos.

Data logistics and AI workloads

As data enters an organization, it becomes subject to a process that IDC describes as data logistics. Using a package-delivery analogy, data logistics is the process of moving data from its origin to its destination, ensuring guaranteed delivery, quality, security, and timeliness. Data logistics provides the foundational framework for AI-RDSI, but AI workloads demand a higher level of rigor than conventional data management systems typically provide **(Figure 1, below).**

Figure 1

AI-enabling data infrastructure



Source: IDC's *AI-Ready Data Storage Infrastructure Special Study,* May 2025

Performance:
End-to-end storage system performance must consider AI workload needs, including data throughput, input/output operations per second (IOPS), latency, the necessary network bandwidth, and the demands of performance-intensive computing. Achieving high throughput may require the use of technologies such as parallel network file systems, as well as flash storage or storage-class memory for IOPS and low latency. Aspects of AI workloads with moderate or lower performance needs may incorporate technologies such as tiered storage, object stores, or hard disk drives for cost efficiency.

Service levels:
Service level requirements go hand in hand with performance, but they are more directed toward data availability. Common service levels include the "nines" of data availability, or total uptime. Five nines (99.999%) uptime will be a common requirement for AI workloads, as downtime will be extremely disruptive.

Data logistics:
Data logistics policy engines assure that data will be delivered to the right place at the right time for AI optimization. This will include location while ensuring adherence to sovereignty requirements.

Data trust:
The desire to feed as much data as possible to AI models for optimized learning must be balanced against the need for data quality. Data trust is core to data quality, as it requires appropriate policies and procedures to reduce data contamination or tampering.

In addition to solid data logistics, the hardware infrastructure must address the needs of AI workloads:

Low latency, high throughput:
AI training and inference workloads place extreme demands on storage systems. Large model training runs require sustained high-throughput data delivery to GPU clusters; any bottleneck at the storage layer can result in expensive GPU idle time. Inference workloads, particularly those serving real-time applications, require sub-millisecond storage latency. Meeting these demands requires state-of-the-art storage architectures rather than retrofitted general-purpose systems.

Tier 0 storage:
Tier 0 refers to the highest-performance tier of the storage hierarchy (typically, non-volatile memory express–based all-flash arrays or storage-class memory) positioned directly in the data path of GPU clusters. Tier 0 storage is characterized by the lowest latency, highest IOPS, and highest cost per terabyte in the storage hierarchy. Its role is to eliminate data starvation in the most performance-critical phases of AI workloads, such as active model training and real-time inference serving.

High-performance file systems (HPFS):
High-performance file systems are designed to manage the massive parallelism and data volumes inherent in AI workloads. Unlike traditional file systems, HPFS architectures distribute metadata and data across multiple nodes, enabling concurrent access by thousands of GPU cores without bottlenecks. Examples include distributed parallel file systems used in high-performance computing environments, which are increasingly being adapted and adopted for enterprise AI infrastructure.

The IDC AI Readiness Index

Based on IDC's methodology of combining survey data with statistical analysis, organizations can be grouped into four tiers of AI data infrastructure readiness. These tiers represent definitive milestones along the maturation journey toward a fully optimized AI-ready data storage infrastructure.

1 Experimenting

Organizations at the Experimenting stage are in the early phases of AI exploration. AI workloads are ad hoc and largely siloed. Data management for AI is predominantly manual, with limited governance and inconsistent quality controls. Storage infrastructure is largely general purpose, with no systematic approach to addressing AI-specific performance requirements. High project failure rates and limited ROI from AI investments characterize these organizations.

2 Practitioners

Practitioner organizations have moved beyond pure experimentation and are beginning to implement repeatable processes. Initial investments in purpose-built AI storage infrastructure are being made. Data governance and quality programs are emerging, though coverage remains incomplete. GPU utilization is improving but is still hampered by storage and data access bottlenecks. AI projects are beginning to reach production but at a lower rate than strategic planning would require.

3 Competent

Competent organizations have established structured and automated approaches to AI data management. Storage infrastructure is architected for AI workloads, with performance tiers aligned to workload requirements. Data governance is systematic and broadly enforced. Data scientists spend the majority of their time on model development rather than waiting on IT resources. AI projects reach production at high rates, and measurable business value is being extracted.

4 Masters

Masters represent the apex of AI data infrastructure maturity. Storage systems are fully optimized for AI workloads: low latency, high throughput, and high availability. Data logistics are automated and governed end to end. Infrastructure is architected with deliberate flexibility to accommodate technological evolution and future AI workload categories. Masters achieve superior ROI on AI investments and translate that operational advantage into sustained competitive differentiation in the marketplace.

Key insight

Not all organizations start at the beginning of this maturity journey, but none start at the end. The IDC AI Readiness Index provides a structured assessment tool that helps IT leaders identify their current position and determine the most impactful next steps toward advancing their readiness.

Future outlook

IDC's global survey of more than 1,300 organizations produced six key findings that illuminate the current state of AI data infrastructure and the gap that most organizations must close to achieve production-grade AI success.

Key finding 1

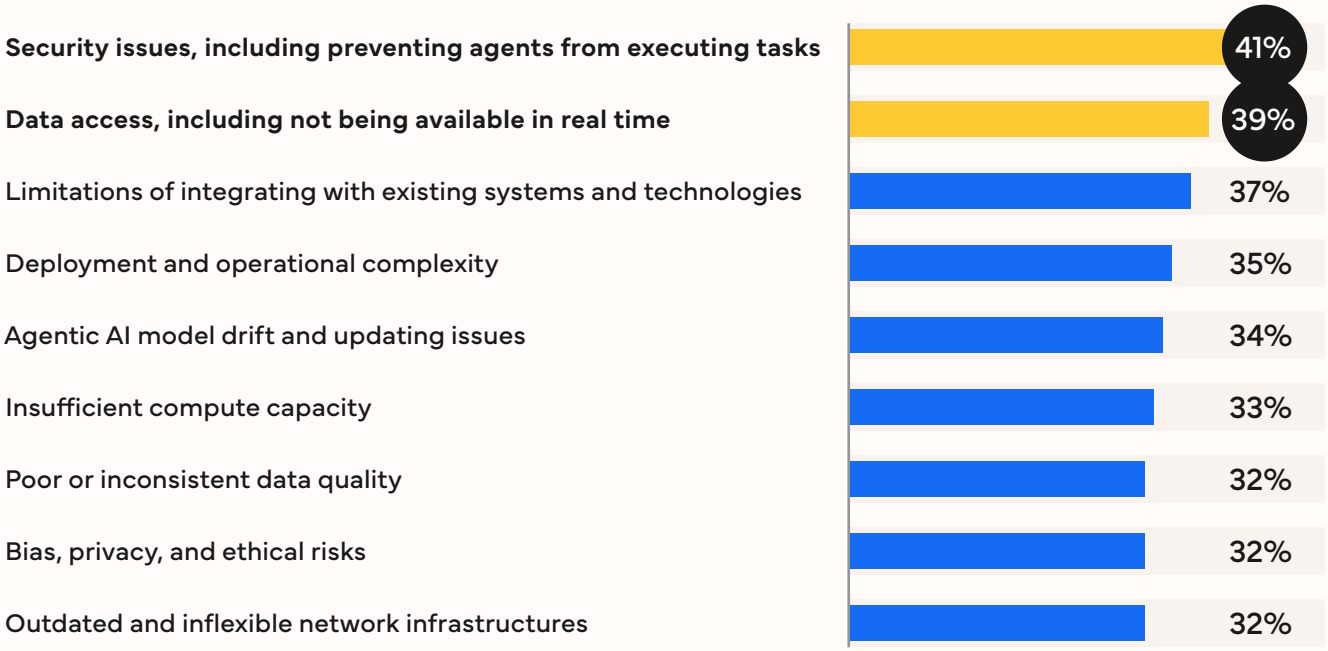
Security and data access are the top 2 challenges to agentic AI deployments (Figure 2, next page).

Figure 2

Top challenges to implementing agentic AI

What are the main challenges in implementing agentic AI in the IT infrastructure of your organization?

(Percentage of respondents)



n = 1,313; Source: IDC's *Global AI Readiness Survey,* March 2026

Agentic AI represents the frontier of enterprise AI deployment: autonomous systems that act on real-time data, execute multi-step workflows, and interact with external services on behalf of users. Its data requirements are correspondingly demanding.

Survey results reveal that data security (41%) is the most commonly cited challenge to implementing agentic AI, followed closely by a lack of adequate data access, including real-time access (39%). These two factors are deeply interrelated: Agentic systems require both broad access to data and strong controls over that access, creating an infrastructure challenge that conventional storage architectures are poorly equipped to address.

41%
cited data security as their top challenge to agentic AI deployment, the single most frequently cited barrier in the survey.

Key finding 2

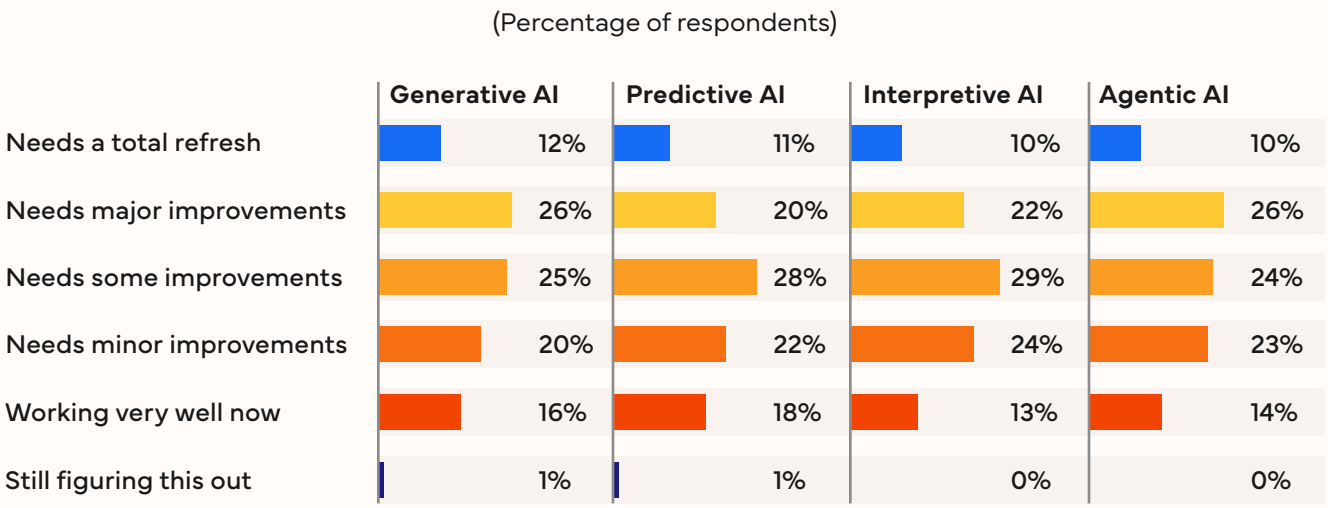
Current storage systems require significant improvement to serve AI workloads (Figure 3, below).

Figure 3

The need for AI storage improvement

What is your assessment of the ability of your organization's data storage infrastructure to support the demands of each of these AI forms?

See the figure data in an accessible table format.



n = 1,313; Source: IDC's *Global AI Readiness Survey,* March 2026

The inadequacy of legacy storage infrastructure for AI workloads is one of the most broadly shared findings in IDC's research. A full 60% of survey respondents indicated that their organization's storage infrastructure requires either significant improvements or a major refresh to adequately support AI workloads.

This finding is consistent across workload types. Whether the workload is GenAI, predictive AI, interpretive AI, or agentic AI, the majority of organizations report that their existing storage systems are not up to the task. The performance, availability, and data management requirements of modern AI workloads have simply outpaced the capabilities of infrastructure designed for earlier generations of computing.

60%
indicated that their storage infrastructure requires significant improvements or a major refresh to support AI workloads.

Key finding 3

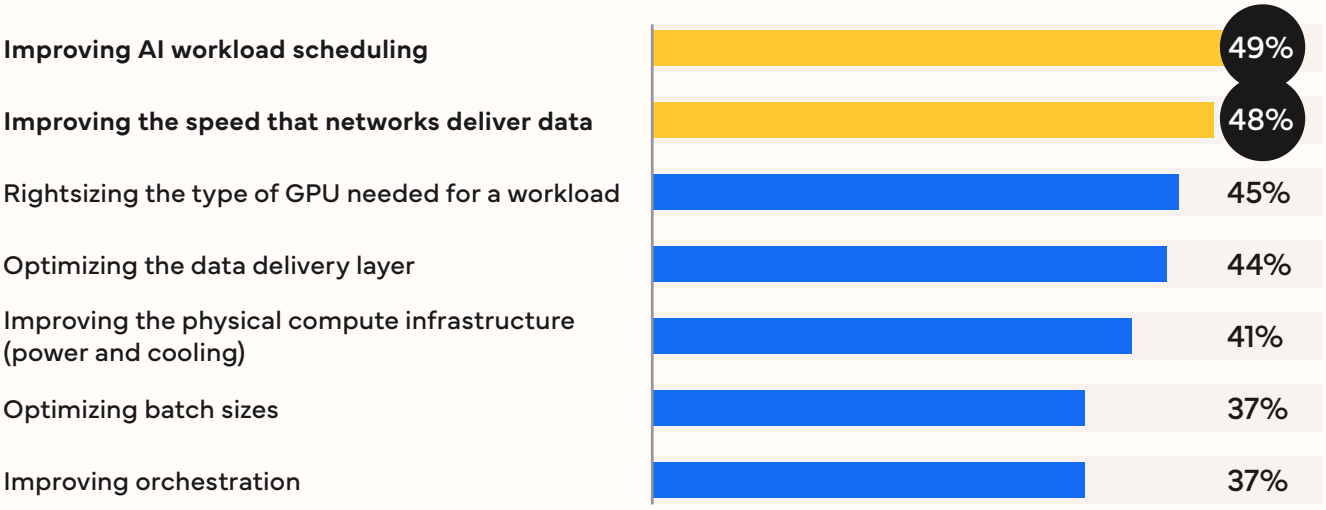
Workload scheduling and network bottlenecks are preventing full GPU utilization (Figure 4, below).

Figure 4

Needed compute infrastructure improvements

What are the main improvements needed for your organization's compute infrastructure to improve GPU utilization?

(Percentage of respondents)



n = 1,313; Source: IDC's *Global AI Readiness Survey,* March 2026

GPUs represent some of the most capital-intensive investments in an AI-capable datacenter. When those GPUs sit idle, waiting for data, the cost of underutilization is significant. Survey data confirms that bottlenecks in data delivery are a pervasive problem: Better workload scheduling (49.4%) and faster data delivery across the network (48.1%) were cited as the top 2 improvements needed to improve GPU utilization.

These findings directly implicate the storage and networking layers as limiting factors in AI infrastructure performance. Workload scheduling improvements are only effective if the underlying data infrastructure can match the cadence of GPU demand. Faster network delivery is only achievable if storage systems can sustain the throughput required. Both findings reinforce the case for well-architected high-performance AI storage infrastructure.

49%
cited a need for better workload scheduling to improve GPU utilization, while 48.1% cited faster data delivery across the network, the findings that indicate both the storage layer and the network layer are key bottlenecks.

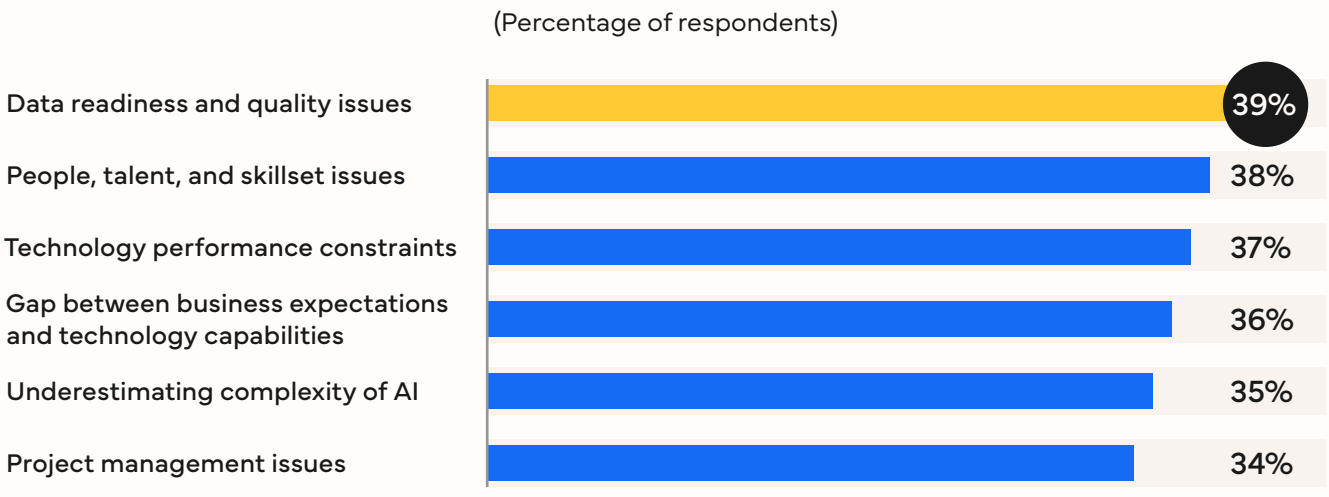
Key finding 4

Data readiness is the primary driver of delays to AI initiatives (Figure 5, below).

Figure 5

Causes of AI initiative delays

What are the main factors driving wait times in AI initiatives in your organization?



n = 1,313; Source: IDC's *Global AI Readiness Survey,* March 2026

When AI projects stall, data issues are most often the reason. Survey respondents identified data readiness and quality as the leading factor delaying AI initiatives, cited by 39.2% of organizations. Talent skills gaps (37.5%) ranked second, but even this factor is partially a manifestation of data challenges, as the shortage of skilled personnel to manage complex data environments compounds the underlying infrastructure problem.

The implication is clear: Organizations that invest in people, processes, and tools to improve data readiness will directly reduce AI project cycle times. Faster time to production translates into faster time to value, a competitive advantage that compounds over time.

39%

identified data readiness and quality as the top factor delaying AI initiatives ahead of all other factors, including talent gaps and budget constraints.

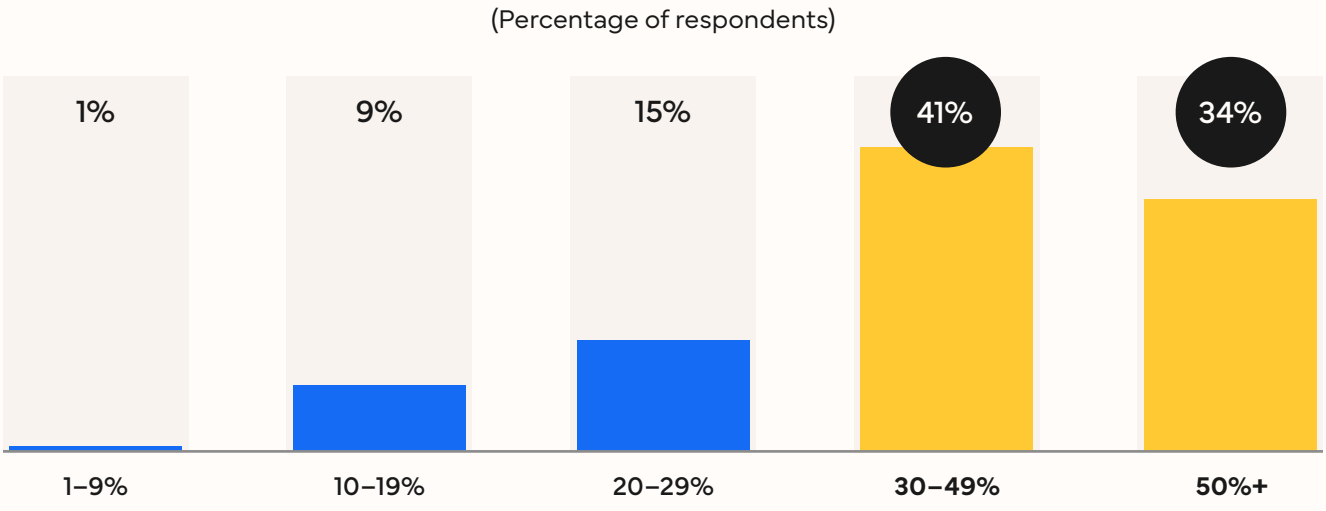
Key finding 5

Data scientists lose significant productive time waiting on IT (Figure 6, below).

Figure 6

Data scientists' time spent waiting on AI requests

What proportion of their time do you believe data scientists spend waiting for IT (including asking for and managing requests for datasets and GPU time) rather than working on models?



n = 1,313; Source: IDC's *Global AI Readiness Survey*, March 2026

Data scientists are among the highest-value and most difficult-to-hire personnel in the technology workforce. Yet current infrastructure conditions mean that a significant portion of their time is spent waiting for datasets to be provisioned, GPU time to be allocated, or IT requests to be fulfilled, among others.

Survey results quantify the scale of this waste: 34.3% of organizations report that their data scientists spend more than half of their time waiting on IT rather than actively working on models. This represents a substantial drag on organizational AI capability and a compelling financial case for infrastructure modernization to automate and accelerate data delivery.

34%
report that data scientists spend more than 50% of their time waiting on IT, rather than doing the model development work they were hired to do.

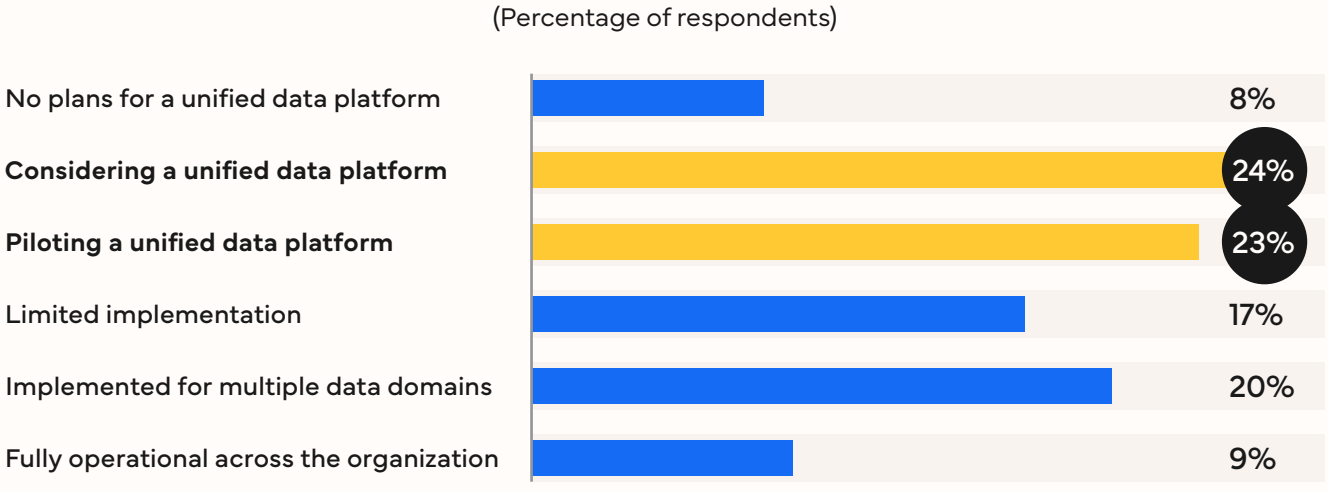
Key finding 6

Unified data platforms are a high-priority infrastructure investment (Figure 7, below).

Figure 7

Plans to deploy unified data platforms

What are your organization's plans for adopting and implementing a unified data platform?



n = 1,313; Source: IDC's *Global AI Readiness Survey,* March 2026

An intelligent data platform — one that provides consistent access to data across silos, environments, and workload types — is viewed as a key enabler of AI-ready infrastructure; 45% of survey respondents are already planning or actively piloting a unified data platform.

Organizations are actively seeking to consolidate fragmented data environments. An intelligent data platform, one that provides consistent access to data across silos, environments, and workload types, is viewed as a key enabler of AI-ready infrastructure; 45% of survey respondents are already planning or actively piloting a unified data platform.

This finding reflects a broader recognition that data fragmentation is not merely an operational inconvenience; it is a structural inhibitor of AI success. Siloed data environments increase the probability of poor data quality, complicate governance, and create access barriers that directly undermine agentic and other real-time AI workloads.

Challenges and opportunities

The AI landscape is evolving at a pace that has few historical precedents. IDC expects significant changes in AI technology, model architectures, and deployment patterns over the next three to five years.

No single storage vendor can meet all of the data requirements for AI workloads. Each vendor will have different strengths and weaknesses that must be evaluated in the context of the IT organization's needs and future road map. Nevertheless, organizations that architect their data infrastructure for agility and flexibility today will be far better positioned to capitalize on those changes, rather than being forced into disruptive and expensive retrofits.

Several strategic considerations merit attention as organizations plan their AI data infrastructure investments:

Planning for change

The data requirements of AI systems are diverse, voluminous, and continuously evolving. No infrastructure investment made today can be assumed sufficient for AI workloads in 2028 or 2030. IT leaders must prioritize architectural flexibility, designing systems that can be extended, scaled, and reconfigured as AI technology matures.

The limits of a single platform

The appeal of a single unified data platform is understandable: It promises to eliminate silos, simplify governance, and reduce management complexity. In practice, however, the data needs of AI systems are so diverse and span so many sources that a true single platform may prove technically impossible or operationally impractical for most organizations.

A more realistic target is meaningful consolidation centered around context-aware storage that reduces or eliminates data silos through well-governed data logistics pipelines. The goal is not uniformity but interoperability, ensuring that data can move efficiently and securely between systems, with consistent quality and provenance, wherever it needs to go.

Governance as a competitive differentiator

As AI systems become more capable and consequential, the governance of the data used to train and operate them becomes a source of competitive differentiation rather than merely a compliance obligation. Organizations that invest in rigorous data governance today will be able to deploy AI systems that are more accurate, more trustworthy, and more defensible to regulators and customers alike. Those that do not will face growing liability exposure as AI regulation matures globally.

Conclusion

AI has become the defining technology investment of this era, but investment alone does not produce results. IDC's research is unambiguous: The difference between AI projects that succeed and those that fail is, to a remarkable degree, a function of data. Data quality, data access, data governance, and the infrastructure that enables all three are the determinants of AI project outcomes.

The evidence from IDC's global survey of more than 1,300 organizations tells a consistent story: Legacy storage infrastructure is broadly failing to meet the demands of modern AI workloads. Bottlenecks at the storage and network layers are undermining GPU utilization, extending project timelines, and wasting the time of the data scientists whom organizations have worked hard to attract and retain.

The path forward is clear, if not always simple. Organizations must:

Assess their current position on the IDC AI Readiness Index and establish a structured maturity road map.

Invest in well-architected, high-performance storage infrastructure capable of meeting the throughput, latency, and availability requirements of AI workloads.

Implement systematic data governance programs that treat data quality, provenance, and trust as first-class operational concerns.

Reduce data silos through strategic consolidation and the deployment of intelligent data platforms where practical.

Architect for flexibility, designing systems that can evolve as AI technology and business requirements change.

Organizations that execute on these priorities will not only achieve higher AI project success rates and stronger ROI but also build a data foundation that sustains competitive advantage for years to come. Those that delay will find the gap widening as their competitors pull ahead.

The question for IT leaders is not whether data infrastructure modernization is necessary. The data makes that case conclusively. The question is how quickly and how deliberately organizations are willing to move and ultimately achieve the capabilities of an AI data Master.

Appendix: Accessible data tables

This appendix provides an accessible version of the data for any complex figures in this document. Click "Return to figure" to get back to the original figure.

Figure 3 accessible data

The need for AI storage improvement

What is your assessment of the ability of your organization's data storage infrastructure to support the demands of each of these AI forms?

Assessment	Generative AI	Predictive AI	Interpretive AI	Agentic AI
Needs a total refresh	12%	11%	10%	10%
Needs major improvements	26%	20%	22%	26%
Needs some improvements	25%	28%	29%	24%
Needs minor improvements	20%	22%	24%	23%
Working very well now	16%	18%	13%	14%
Still figuring this out	1%	1%	0%	0%

n = 1,313; Source: IDC's *Global AI Readiness Survey,* March 2026

Return to figure

About the IDC analyst

Phil Goodwin

Research Vice President,
Infrastructure Software Platforms,
Worldwide Infrastructure Research, IDC

Phil Goodwin is research vice president within IDC's Worldwide Infrastructure Research organization and global research lead for the Infrastructure Software Platforms practice. He leads a team of analysts that provide detailed insights and analyses on evolving infrastructure software trends, vendor performance, and the impact of new technology adoption. Goodwin's own research focuses on multi-cloud data management, data logistics, on-premises and cloud-based data protection as a service, cyberprotection and recovery, and recovery orchestration. He takes a holistic view of these markets and covers risk analysis, service-level requirements, and cost/benefit calculations in his research. Goodwin also contributes regularly to IDC's CIO Advisory practice.

More about Phil Goodwin →

Message from the sponsor

AI ambition has outpaced AI delivery. Organizations are pursuing faster decisions, better customer experiences, and new sources of competitive advantage — yet most projects stall before reaching production. The reason isn't ambition. It isn't compute. Most of the time it's data.

The findings from IDC's research confirm what we see across the enterprise landscape: 54% of AI proof-of-concept projects never reach production, and 60% of organizations say their storage is not ready for AI workloads.

Everpure gives infrastructure teams one platform for the full AI lifecycle — training, inference, and agents — with the operational simplicity, security, and governance required to turn AI ambition into production results.

Go from pilot to production with data that's ready for AI.

Discover the platform

See how your AI progress stacks up

IDC Custom Solutions

IDC Custom Solutions produced this publication. The opinion, analysis, and research results presented herein are drawn from more detailed research and analysis that IDC independently conducted and published, unless specific vendor sponsorship is noted. IDC Custom Solutions makes IDC content available in a wide range of formats for distribution by various companies.

IDC.com | IDC on LinkedIn | IDC Blog

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications, and consumer technology markets. With more than 1,300 analysts worldwide, IDC offers global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries. IDC's analysis and insight helps IT professionals, business executives, and the investment community to make fact-based technology decisions and to achieve their key business objectives.