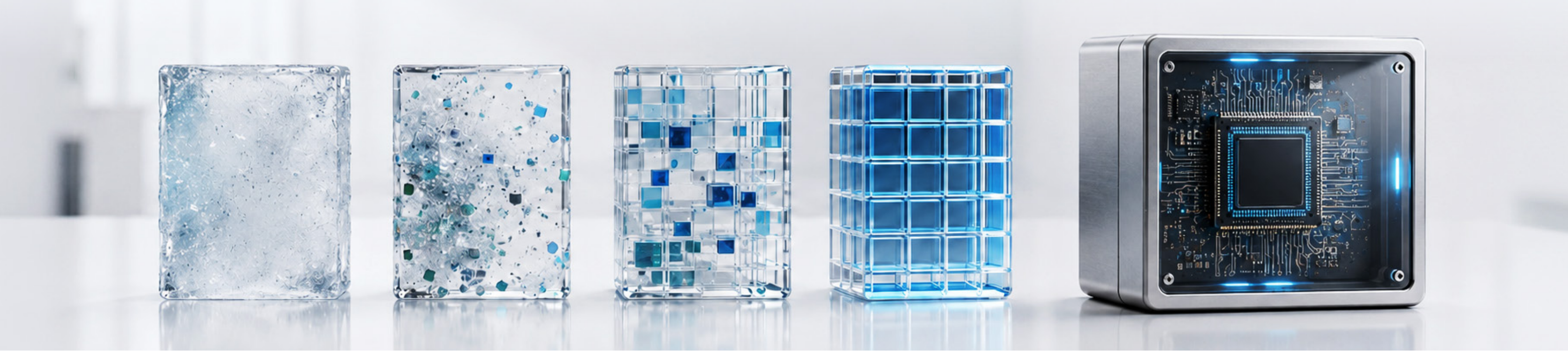
CloudKinetics

Why Data Readiness Is A Big AI Success Factor



Key Takeaways

- Data readiness is the deciding factor in whether enterprise AI use cases succeed or fail. It's what allows AI to function, the way senses allow a mind to perceive. AI leans on data to understand, learns from it and acts on it, making data the fuel that drives AI.
- Five conditions define data readiness for AI: Data quality, cross-system accessibility, governance with clear lineage, structured metadata and real-time availability.
- Gartner projects that 60% of AI initiatives lacking AI-ready data will be abandoned. Weak data foundations account for the majority of stalled enterprise AI projects.
- Four practices are consistently associated with data-ready organizations: A unified data platform, data governance and controls, automated cloud-native pipelines and cross-functional data ownership.
- Data readiness functions as an ongoing operating capability. Continuous governance and infrastructure processes support the adoption of new AI use cases as they emerge.

According to a McKinsey report, only 7% of companies have fully scaled AI across their organizations, a gap that traces back less to model choice and more to the data feeding those models.

"Data readiness" has moved from a background IT concern to a factor that directly determines whether an AI initiative delivers business value. The differentiator is whether the underlying data can support the model reliably, consistently and at the pace AI systems require.

This shift has implications for how enterprise leaders evaluate AI investment. Alongside model selection and use case prioritization, data infrastructure readiness has become an important consideration that will determine the course and future of enterprise AI.

> "AI without data is a mind without memory – capable of thought, but incapable of experience. AI turns data into decisions, but only if the data was worth deciding on in the first place. The organizations that win with AI won't be the ones with the best models. They'll be the ones with the best-kept data."
>
> **Punit Chheda, VP & Lead Architect – Data & Applied AI Practice**

What data readiness means for enterprise AI systems

Data readiness refers to five specific conditions an organization's data must meet before AI systems can use it reliably:

Quality
Data must be accurate, complete and free of duplication.

Metadata management
Labels and structure that make data usable rather than just present.

Accessibility
Information locked in siloed systems cannot feed AI models regardless of its quality.

Real-time availability
AI systems designed for current decision-making require data that is current.

Governance and lineage
Tracking where data originates, who owns it and how it changes over time.

These conditions carry more weight today than they did during earlier phases of enterprise AI adoption. Predictive analytics tools could often work with historical, batch-processed data. Agentic and GenAI systems operate differently. They require continuous access to current information, and they need to trace that information back to a verified source before a business can act on their output.

Gartner projects that 60% of AI initiatives lacking AI-ready data will be abandoned soon. The gap between AI ambition and data infrastructure is measurable.

The cost of poor data readiness in AI projects

Gartner's Hype Cycle for Generative AI found that a significant share of generative AI projects will exceed their budgeted costs due to weak architectural choices and limited operational experience. Some organizations are abandoning custom model efforts entirely due to cost and technical debt.

> The costs associated with inadequate data readiness show up across several dimensions of project timelines, infrastructure spend and compliance exposure.

Compliance exposure adds another layer. Data without clear lineage or governance creates audit risk under frameworks such as the EU AI Act, particularly for systems classified as high-risk. Organizations without documented data ownership face longer remediation timelines when regulators request that documentation.

Across these dimensions, a common pattern is that data infrastructure gaps identified late in a project tend to carry higher remediation costs than those addressed before deployment begins.

4 practices that support reliable AI deployment

Organizations that succeed with AI treat data infrastructure as a foundation, built before deployment begins. Four practices are associated with organizations that have successfully moved AI initiatives into production

- **A unified data platform:** Often built on a data fabric or lakehouse architecture that connects information across departments into a single accessible layer. This gives AI systems a more complete and consistent view of the data they need to operate reliably
- **Governance frameworks:** Assign ownership at the level of individual data assets, not only at the department level. Access rules, quality standards and change tracking are defined for each asset and checked on an ongoing basis. This differs from governance models built around periodic review cycles, which were designed for analytical rather than operational AI use. AI systems that require current, verified information benefit from governance processes that operate continuously.
- **Data pipelines:** Automate the work of collecting, cleaning and moving data between systems. Automation reduces manual integration effort, which becomes a scaling constraint as AI systems move from pilot deployments into full production use.
- **Ownership:** Sits jointly across business and technical teams, rather than resting solely with IT. Business teams define what the data needs to support, while technical teams build and maintain the infrastructure. This shared model prevents the common gap where IT builds a system that does not match how the business intends to use it.

Unified data platform

A single accessible layer across departments

Governance framework

Asset-level ownership and continuous quality checks

Cloud native pinelines

Automated ingestion and transformation of data

Cross functional ownership

Shared accountability between business and IT teams

Snowflake's research on enterprise AI adoption found that early adopters report positive returns, with approximately $1.49 in value generated for every $1 invested in AI. The pattern holds across industries, with readiness work completed before deployment reducing the volume of rework required after.

How to assess data readiness before scaling AI

Data readiness requirements vary by use cases. Before assessing readiness, organizations need to identify what data is required for AI systems. The type, volume and frequency of data requirements vary from system to system. Planning readiness around the specific use case ensures the right kind of data is prioritized.

- **Data quality:** Completeness, accuracy and consistency across relevant datasets. Data should be processed through defined stages from raw ingestion to gold layers. This is largely done through the Medallion architecture, where data is organized in a data lake or lakehouse. These are often paired with open table formats like Apache Iceberg, keeping large-scale datasets query-ready for AI systems.
- **Accessibility:** Whether required data can be reached across the systems involved or remains confined to a single department or legacy application. This is creating a shift from traditional business intelligence (BI) to AI-driven BI. Instead of static dashboards and scheduled reports, users prompt AI directly in natural language. This is only possible when the data is processed, transformed and made available for AI to query.
- **Governance maturity:** Whether ownership, access permissions and audit trails exist for each relevant data asset. This begins by classifying and tagging the datasets, which determines the access permissions and audit trails for AI systems.
- **Metadata coverage:** Whether data has sufficient context and structure for AI systems to interpret correctly. This includes clear labelling, defined data types and the relationship between datasets, as this gives the context for AI to distinguish.
- **Real-time availability:** Whether data refreshes at a pace that matches the AI system's intended use. The weight given to real-time availability should scale with how time-sensitive an AI use case is.

Gartner recommends building on existing data management practices by adding AI-specific capabilities such as vector data stores and retrieval-augmented generation (RAG) support incrementally, rather than building new data infrastructure from scratch for each use case.

Data readiness is an ongoing priority

For organizations building data readiness capabilities, the work spans several distinct areas, each of which can be addressed independently or in combination depending on what the current infrastructure requires.

Data platform modernization supports the move from fragmented, on-premises systems toward unified, cloud-native architectures designed to handle AI workloads at scale. Governance-as-a-service introduces structured ownership, access control and audit-ready documentation at the asset level.

Data cost management, closely aligned with FinOps practices, tracks and controls the infrastructure spend associated with storing, processing and moving large data volumes. This keeps AI initiatives within budget as they scale. Delivery models that span multiple regions allow organizations with distributed operations to apply consistent data standards across locations, rather than managing separate approaches per market.

> The success of AI implementation depends on how fast quality data is available and accessible for users and AI to consume.

Frequently asked questions (FAQs)

What is data readiness and why is it important?

Data readiness is the extent to which an organization's data is prepared for reliable AI use. It depends on factors such as data quality, accessibility, governance, metadata and timely availability. When these elements are in place, AI systems can access trusted and well-structured data. This helps organizations build more reliable and maintainable AI applications.

How can organizations measure their current level of data readiness?

Organizations can assess readiness against defined checkpoints of data quality, accessibility, governance maturity, metadata coverage and real-time availability. Scoring each checkpoint for a specific use case gives a clearer picture than a general audit.

What are specific strategies to overcome siloed data for AI initiatives?

A unified data platform, built on a data fabric or lakehouse architecture, connects data across departments into one accessible layer. Cloud-native pipelines automate the movement of data between systems, reducing the manual integration work that keeps data siloed. Assigning cross-functional ownership, rather than leaving data management solely with IT, also helps break down silos

---

What are common pitfalls when implementing enterprise data governance for AI?

Governance frameworks that operate on quarterly or annual review cycles fall behind AI systems that need data quality checked continuously. Assigning governance responsibility to a single team, without cross-functional input, often results in rules that do not match how data is actually used. Skipping asset-level ownership, and instead applying governance broadly at the department level, leaves specific data assets without a clear owner.

How can organizations get support in building data readiness?

Managed services can address specific gaps, including data platform modernization, governance-as-a-service and data cost management aligned with FinOps practices. These options allow organizations to build a data foundation for AI without developing every capability internally, particularly useful for those with distributed operations across multiple regions or limited in-house data engineering capacity.

From data platform modernization to cost management, we work to build the data foundation your AI initiatives depend on. If you are ready to assess your enterprise's data readiness, connect with us!

www.cloud-kinetics.com | contactus@cloud-kinetics.com

Get in touch with us