LUMO INSIGHTS

Thursday, 14 May 2026

5% of enterprises have AI-Ready Data - what sets them apart

Executive Summary

This week’s developments highlight a sobering truth: the biggest obstacles to AI success aren’t algorithms - they’re data challenges. New surveys and strategic moves by tech leaders reveal an industry-wide push to fix data quality, integration, and governance issues that have been quietly holding back AI initiatives. The organisations that master their data foundations are rapidly pulling ahead, while those that don’t risk seeing their AI ambitions stall out.

The Data readiness divide

A wave of fresh research reveals a growing gap between AI ambition and data reality in today’s enterprises. While nearly every company is experimenting with artificial intelligence, few have the data foundations to show for it. According to a new Dun & Bradstreet “AI Momentum” survey, 97% of organisations report active AI initiatives underway, yet only 5% say their data is fully ready to support AI at enterprise scale (www.cio.com [1]). This immense disparity underscores that flashy algorithms alone won’t carry a business to success. It’s the unglamorous work of preparing data - cleaning it, integrating it across silos, and enforcing quality and governance - that increasingly determines which AI projects thrive.

Challenges around data readiness have only deepened over the past year. In the D&B study, 50% of companies cited inadequate access to data as a major barrier to AI adoption; 44% flagged privacy and compliance issues; 40% pointed to poor data quality; and 38% reported a lack of integration across internal systems (www.cio.com [2]). With so many enterprises still struggling to get basic data infrastructure in order, these problems have become serious bottlenecks slowing AI programmes.

As D&B’s Chief Strategy Officer, Cayetano Gea-Carrasco, observed, the question for companies now is "no longer whether organisations are experimenting with AI … it’s whether they have the data and infrastructure required to deploy AI reliably at enterprise scale" (www.cio.com [3]). In other words, without a robust data architecture, even the most impressive pilot will remain difficult to scale into a mission-critical, AI-powered workflow.

On the bright side, a small minority of AI leaders are showing what works. These front-runners make sure their data is high-quality, timely, and unified across the organisation (www.cio.com [4]). They invest in consistent data definitions and modern pipelines to break down silos, linking customer, operational, and third-party information on integrated platforms. With strong governance and metadata management in place, both employees and AI systems can easily find and trust the data they need. These data-savvy leaders are already enjoying outsized AI-driven returns, while others remain mired in one-off pilots.

References:

[1] www.cio.com — https://www.cio.com/article/4170978/nearly-every-enterprise-is-investing-in-ai-but-only-5-say-their-data-is-ready.html#:~:text=a%20new%20AI%20Momentum%20Survey,critical
[2] www.cio.com — https://www.cio.com/article/4170978/nearly-every-enterprise-is-investing-in-ai-but-only-5-say-their-data-is-ready.html#:~:text=data%20readiness%20are%20%E2%80%9Ceven%20more,infrastructure%20required%20to%20deploy%20AI
[3] www.cio.com — https://www.cio.com/article/4170978/nearly-every-enterprise-is-investing-in-ai-but-only-5-say-their-data-is-ready.html#:~:text=37,into%20production%20workflows%2C%20where%20accuracy
[4] www.cio.com — https://www.cio.com/article/4170978/nearly-every-enterprise-is-investing-in-ai-but-only-5-say-their-data-is-ready.html#:~:text=industries%20like%20banking%2C%20insurance%2C%20healthcare%2C,This%20includes%20areas%20like%20sales

When bad Data derails good AI

Companies are also learning that neglecting data governance can turn even promising AI projects into expensive failures. Many organisations have rushed to appoint chief AI officers and establish AI ethics committees, but in doing so they often neglect "the state of their own data" (newsroom.ibm.com [1]). If the information feeding machine-learning models is incomplete, outdated, or poorly managed, the results will be flawed - or even harmful - no matter how advanced the algorithms.

One high-profile example illustrates the stakes. Starbucks recently deployed an AI-driven inventory management system to automate stock checks in its coffee stores. But the system was fed bad data, and it ended up making inaccurate reordering decisions that led to unused stock in some locations while others ran out of items - ultimately causing wasted inventory and lost sales (newsroom.ibm.com [2]). Starbucks' experience is no outlier. In a recent Gartner survey of IT operations leaders, 38% reported that poor data quality or limited data availability was a direct cause of an AI project failure in their organisation (www.rand.org [3]). The message is clear: bad data can quickly sink good AI, leading to tangible financial and reputational damage.

Data biases present another serious hazard. AI models trained on skewed or incomplete datasets may end up amplifying discrimination and making unjust decisions. For example, a 2025 study reported that a hospital's experimental diagnostic AI - trained on historical emergency department records - was more likely to recommend aggressive treatments for Black, LGBTQ+ and homeless patients than for other groups with the same symptoms (newsroom.ibm.com [4]). Such biased outcomes not only raise ethical and legal issues, but also undermine trust with customers and regulators. And beyond bias, poor data management can erode transparency and accountability. If companies lack clear data provenance and documentation, they risk accumulating "documentation debt" - making it hard to explain how an AI arrived at a decision when auditors or authorities ask. Indeed, repurposing data without a clear lawful basis can breach major privacy laws like GDPR (newsroom.ibm.com [5]) - a costly mistake as regulators ramp up enforcement.

The upshot is that AI readiness must start with getting data right. Before scaling any AI initiative, leading enterprises are conducting rigorous data-quality audits, establishing strong governance policies, and assigning clear ownership for data stewardship. Industry advisors emphasise that successful AI governance truly 'begins with data governance' (newsroom.ibm.com [6]). By fixing data issues at the source and maintaining high standards, companies can prevent small data errors or biases from snowballing into big AI failures.

References:

[1] newsroom.ibm.com — https://newsroom.ibm.com/2025-11-13-ibm-study-chief-data-officers-redefine-strategies-as-ai-ambitions-outpace-readiness#:~:text=Nov,highlights%20a
[2] newsroom.ibm.com — https://newsroom.ibm.com/2025-11-13-ibm-study-chief-data-officers-redefine-strategies-as-ai-ambitions-outpace-readiness#:~:text=high%20amid%20AI,of%20surveyed%20leaders%20have%20started
[3] www.rand.org — https://www.rand.org/pubs/research_reports/RRA2680-1.html#:~:text=to%20succeed
[4] newsroom.ibm.com — https://newsroom.ibm.com/2025-11-13-ibm-study-chief-data-officers-redefine-strategies-as-ai-ambitions-outpace-readiness#:~:text=developing%20diverse%20datasets%20to%20train,their%20organization%20relying%20on%20outcomes
[5] newsroom.ibm.com — https://newsroom.ibm.com/2025-11-13-ibm-study-chief-data-officers-redefine-strategies-as-ai-ambitions-outpace-readiness#:~:text=full%20study%2C%20visit%3A%20https%3A%2F%2Fwww.ibm.com%2Fthought,AI%20Officer%20and%2

0other%20senior
[6] newsroom.ibm.com — https://newsroom.ibm.com/2025-11-13-ibm-study-chief-data-officers-redefine-strategies-as-ai-ambitions-outpace-readiness#:~:text=agree%20they%20can%20clearly%20convey,AI%20ambitions%20remain

Reengineering Data architecture for AI

Amid this intensified focus on data readiness, organisations are rethinking their IT architecture to feed AI's voracious data appetite. The past week saw a major bet on open data infrastructure: SAP announced plans to acquire Dremio, a data lakehouse platform, as part of its strategy to remove silos between SAP's own applications and external sources. SAP's CTO, Philipp Herzig, explained: "Enterprise AI projects often don't stall because of modelling limitations - they stall because the data isn't ready for AI agents" (blobstreaming.org [1]). In other words, fragmented, locked-up data has become a primary blocker to AI value. By bringing Dremio's technology into its Business Data Cloud, SAP will make its customers' data landscape Apache Iceberg-enabled - an open standard that lets different tools share data easily - and provide a universal metadata catalogue to preserve business context and governance across sources (blobstreaming.org [2]). The goal is to give enterprises a single, high-performance platform for both SAP and non-SAP data, eliminating tedious data wrangling so AI models can deliver insights faster.

Another trend is the rise of vector databases - specialised systems for storing and searching the numerical "embeddings" that represent text, images, and other data for AI. As companies deploy generative and 'agentic' AI applications, they need robust, real-time retrieval of knowledge from their vast proprietary datasets. Vector databases now serve as a crucial bridge between large language models and an enterprise's private information, enabling AI to retrieve up-to-date, relevant facts in context. At the recent Data Summit 2026, experts underscored that AI is only as smart as the data it can access - one technologist noted that the 'retrieval layer is the one thing you fully control' when deploying AI agents (www.gartner.com [3]). Reflecting this need, industry forecasts project the vector database market will swell from roughly $1.6 billion in 2023 to over $10 billion by 2032 (www.cio.com [4]) as organisations invest in new retrieval infrastructure to deliver the right data to the right AI systems at the right time.

Enterprises are also adapting their data architectures to satisfy regulatory demands and maintain control over information. In highly regulated sectors like finance, healthcare, and government, data sovereignty and latency requirements are shaping how AI is deployed. For example, Japan's SoftBank has built a new 'sovereign' AI cloud with Oracle that ensures all sensitive data stays within local data centres (news.sap.com [5]). By embracing such sovereign cloud models and hybrid deployments, organisations can harness advanced AI capabilities while complying with data residency laws and keeping critical information under strict governance. From adopting open standards and real-time pipelines to unifying data across multi-cloud environments, the architecture decisions made now will determine which companies can scale AI safely and effectively.

References:

[1] blobstreaming.org — https://blobstreaming.org/notes/vector-databases-in-2026/#:~:text=Vector%20Databases%20in%202026%3A%20Trends,tradeoffs%2C%20and%20the%20decision%20framework
[2] blobstreaming.org — https://blobstreaming.org/notes/vector-databases-in-2026/#:~:text=Vector%20Databases%20in%202026%3A%20Trends,tradeoffs%2C%20and%20the%20decision%20framework
[3] www.gartner.com — https://www.gartner.com/en/newsroom/press-releases/2026-04-07-gartner-says-artificial-intelligence-projects-in-infrastructure-and-operations-stall-ahead-of-meaningful-roi-returns#:~:text=transformation%20by%20integrating%20AI%2C%20optimizing,informed%20decisions%20and%20unlock%20the
[4] www.cio.com — https://www.cio.com/article/4166881/sap-to-acquire-data-lakehouse-vendor-dremio.html#:~:text=Databricks%2C%20but%20analysts%20point%20to,business%20context%20that%20makes%20it
[5] news.sap.com — https://news.sap.com/2026/05/sap-to-acquire-dremio-unify-sap-and-non-sap-data-power-agentic-ai/#:~:text=for%20real,meaning%2C%20relationships%2C%20access%20rights%20and

Data as the new competitive moat

The clearest message from this week's developments is that data itself is emerging as the ultimate competitive advantage in the AI era. According to a recent IBM global study, 78% of Chief Data Officers say that leveraging proprietary data is now their top strategy for market differentiation . In a world where cutting-edge AI models are increasingly accessible to all, companies are realising that their most valuable asset is the unique data - about their customers, operations, and domain - that nobody else possesses. By doubling down on data collection, curation, and governance, they aim to build durable "data moats" that rivals cannot easily cross.

A striking case in point comes from Thomson Reuters, which turned its trove of legal and tax documents into a source of competitive strength. The company's stock jumped more than 11% after it announced that its CoCounsel AI assistant had surpassed 1 million users (news.sap.com [1]). This surge defied a broader tech downturn and reflected investor confidence in Thomson Reuters' unparalleled data assets. Executives emphasised that while general-purpose AI platforms may be powerful, they lack access to Thomson's 150-year archive of authoritative legal information (muawia.com [2]). In effect, Thomson Reuters has built a truly defensible data moat - transforming centuries of proprietary content into AI-powered services that competitors cannot easily replicate.

Business leaders are increasingly taking this lesson to heart. Industry observers note that organisations winning with AI treat data as a strategic resource, applying the same rigour to data governance and quality as they do to financial controls . Those that move early to modernise and protect their data foundations will compound their advantage for years to come, whereas companies that wait too long to act are likely to find the cost of catching up grows exponentially . The race is on - turning enterprise data into AI-driven insight and innovation is fast becoming the defining competitive battleground of this decade.

References:

[1] news.sap.com — https://news.sap.com/2026/05/sap-to-acquire-dremio-unify-sap-and-non-sap-data-power-agentic-ai/#:~:text=enterprise%20lakehouse%20that%20unifies%20SAP,agentic%20AI%20at%20enterprise%20scale
[2] muawia.com — https://muawia.com/databricks-unveils-ai-lakehouse-for-enterprise-dat-2026-03-31/#:~:text=Databricks%20Unveils%20AI%20Lakehouse%20for,work%20smarter%20and%20automate%20workflows

Key Statistics

- 97% of organisations have active AI initiatives, but only 5% say their data is fully ready for AI at scale ([www.cio.com](https://www.cio.com/article/4170978/nearly-every-enterprise-is-investing-in-ai-but-only-5-say-their-data-is-ready.html#:~:text=a%20new%20AI%20Momentum%20Survey,critical)).
- 38% of IT leaders said poor data quality or limited data availability was a direct cause of an AI project failure ([www.rand.org](https://www.rand.org/pubs/research_reports/RRA2680-1.html#:~:text=to%20succeed)).
- Gartner analysts predict that by 2026, 60% of all AI projects will be cancelled due to inadequate data foundations ([www.dw.com](https://www.dw.com/en/eu-plans-to-ease-gdpr-laws-and-ai-constraints-in-major-shift/a-74792773#:~:text=Data%20privacy%20laws%20and%20protections,released%20by%20the%20EU%20on)).
- 78% of surveyed Chief Data Officers cite leveraging proprietary data as a top strategic objective for market differentiation .

KEY TAKEAWAY

For business leaders, the message is clear: companies winning with AI treat data as a strategic asset. Now is the time to invest in unified, high-quality, well-governed data foundations to unlock scalable AI value and build a lasting competitive edge.

Sources

Nearly every enterprise is investing in AI, but only 5% say their data is ready – CIO (May 13, 2026)
https://www.cio.com/article/4170978/nearly-every-enterprise-is-investing-in-ai-but-only-5-say-their-data-is-ready.html

A.I. Adoption Is Surging. Data Governance Is Not Keeping Up. – Observer (May 11, 2026)
https://observer.com/2026/05/ai-adoption-data-governance-enterprise-risk/

SAP to Buy Dremio and Prior Labs to 'Lead' Agentic AI Models – CRN (May 4, 2026)
https://www.crn.com/news/ai/2026/sap-to-buy-dremio-and-prior-labs-to-lead-agentic-and-ai-models

Thomson Reuters shares rally after CoCounsel AI tool draws 1 million users – Reuters (Feb 24, 2026)
https://money.usnews.com/investing/news/articles/2026-02-24/thomson-reuters-shares-rally-after-cocounsel-ai-tool-draws-1-million-users

What's New in AI? May 2026 Edition (SoftBank's Sovereign Cloud) – Oracle AI Blog (May 8, 2026)
https://blogs.oracle.com/ai-and-datascience/whats-new-in-ai-may-2026

Gartner Q&A: AI Projects in I&O Stall Ahead of Meaningful ROI Returns – Gartner Newsroom (Apr 7, 2026)
https://www.gartner.com/en/newsroom/press-releases/2026-04-07-gartner-says-artificial-intelligence-projects-in-infrastructure-and-operations-stall-ahead-of-meaningful-roi-returns

Data Summit 2026: Trust in Data Key to AI Scaling – EnterpriseAIWorld (May 6, 2026)
https://www.enterpriseaiworld.com/Articles/News/News/Data-Summit-2026-Day-1-Keynotes-Offer-a-New-Way-to-See-Data-Through-the-Lens-of-AI-174686.aspx

Vector Database Market to Reach $10.6 Billion by 2032 – Yahoo Finance (Mar 7, 2025)
https://finance.yahoo.com/news/vector-database-market-reach-usd-150000060.html

IBM Study: Chief Data Officers Redefine Strategies as AI Ambitions Outpace Readiness – IBM Newsroom (Nov 13, 2025)
https://newsroom.ibm.com/2025-11-13-ibm-study-chief-data-officers-redefine-strategies-as-ai-ambitions-outpace-readiness

The New Moat: Why Proprietary Data Is Your Only Durable Competitive Advantage in AI – AI Ireland (Mar 25, 2026)
https://aiireland.ie/2026/03/25/the-new-moat-why-proprietary-data-is-your-only-durable-competitive-advantage-in-ai/