LUMO INSIGHTS

Monday, 15 June 2026

Data strategy emerges as the make-or-break factor for AI success

Executive Summary

New developments in the past week show that effective data strategy is now the decisive factor in enterprise AI success. Many AI initiatives are struggling not because of model limitations, but due to data quality, integration, and governance issues. Forward-looking companies - and regulators worldwide - are focusing on modern data architectures, proprietary data assets, and stricter controls as key to unlocking AI's full business value.

Modern Data infrastructure: the key to AI leadership

In the race for AI-driven performance, a new gulf is opening between companies that have built strong data foundations and those that haven't. A new global survey by BCG reveals that only a small elite - about 5% of firms - are "future-built" for AI, capturing outsized value from their investments, while roughly 60% of organisations report only minimal gains despite substantial AI spending (www.bcg.com [1]). These AI leaders are pulling far ahead - achieving five times the revenue increases and three times the cost reductions from their AI initiatives compared to lagging competitors (www.bcg.com [2]).

What are these leaders doing differently? They are investing in modern, scalable data infrastructures that allow AI to be deployed enterprise-wide without hitting performance bottlenecks. For example, more than half of "AI leader" companies (52%) have adopted a hybrid cloud data strategy, versus just 35% of other organisations (redmondmag.com [3]). This kind of blended approach lets them run AI workloads wherever it makes the most sense, combining the flexibility of public cloud with the control and compliance benefits of private on-premises systems (redmondmag.com [4]). Successful AI-first enterprises are also increasingly relying on cohesive data platforms that span both cloud and on-site environments, breaking down silos and enabling advanced AI applications to scale seamlessly across the organisation (redmondmag.com [5]).

Top executives are recognising the competitive necessity of these data investments. In a recent Cisco-sponsored survey, 97% of IT and business leaders said they plan to expand AI use, but 74% admitted that outdated data and network infrastructure is already holding back their growth (newsroom.cisco.com [6]). Nearly all (96%) also believe that modernising their technology - often in partnership with trusted data platform providers - is critical for future success in the AI era (newsroom.cisco.com [7]). The takeaway is clear: to join the ranks of AI leaders, companies must double down on building future-ready data architectures today or risk falling behind.

References:

[1] www.bcg.com — https://www.bcg.com/publications/2025/are-you-generating-value-from-ai-the-widening-gap#:~:text=built%20companies%20achieve%20five%20times,cost%20gains%20despite%20substantial%20investment

[2] www.bcg.com — https://www.bcg.com/publications/2025/are-you-generating-value-from-ai-the-widening-gap#:~:text=built%20companies%20achieve%20five%20times,cost%20gains%20despite%20substantial%20investment

[3] redmondmag.com — https://redmondmag.com/articles/2025/06/20/new-reports-identify-traits-of-enterprise-ai-leaders-and-laggards.aspx#:~:text=cloud%20computing,premises%20environments%20to%20support
[4] redmondmag.com — https://redmondmag.com/articles/2025/06/20/new-reports-identify-traits-of-enterprise-ai-leaders-and-laggards.aspx#:~:text=cloud%20computing,premises%20environments%20to%20support
[5] redmondmag.com — https://redmondmag.com/articles/2025/06/20/new-reports-identify-traits-of-enterprise-ai-leaders-and-laggards.aspx#:~:text=combining%20the%20flexibility%20of%20public,approaches%20of%20the%20studies%20are
[6] newsroom.cisco.com — https://newsroom.cisco.com/c/r/newsroom/en/us/a/y2025/m06/cisco-research-a-major-infrastructure-shift-is-underway-ai-could-double-the-strain-or-solve-it.html#:~:text=Cisco%E2%80%99s%20recent%20research%20shows%20CEOs,based%20on%20a%20survey%20of
[7] newsroom.cisco.com — https://newsroom.cisco.com/c/r/newsroom/en/us/a/y2025/m06/cisco-research-a-major-infrastructure-shift-is-underway-ai-could-double-the-strain-or-solve-it.html#:~:text=on%20their%20CIO%20or%20CTO,key%20findings%20from%20Cisco%E2%80%99s%20Networking

Data quality & governance: AI's hidden bottleneck

The biggest obstacles to AI success are often not technical at all - they boil down to data quality, organisation, and governance. Multiple studies confirm that most AI projects don't falter because of model shortcomings; they fail because the data behind the AI is incomplete, unclean, or poorly managed (www.bain.com [1]). As one blunt industry analysis put it, "data quality kills more projects than any algorithm flaw" (www.valuebound.com [2]) - a hidden crisis that is coming to the forefront as AI initiatives scale up.

For example, an MIT Sloan study reports an astonishing 95% of AI projects fail to deliver on their promises (www.forbes.com [3]). The primary reason is that only a tiny fraction of a typical enterprise's information is truly "AI-ready" - meaning organised, accurate, and up-to-date enough for effective model training (www.forbes.com [4]). One Gartner analysis estimated that poor data quality costs businesses about $12.9 million per year and is responsible for 40% of all failed business initiatives (www.forbes.com [5]). In healthcare, an estimated 72% of AI projects have failed to achieve their objectives due to weak data governance - medical data is trapped in silos and incompatible formats, so algorithms simply can't get reliable inputs (www.healthittek.com [6]) (www.healthittek.com [7]).

On the flip side, organisations that have their data house in order are seeing AI projects thrive. Some of the most impressive AI breakthroughs today are coming from using better data rather than more complex models. In one case, a small analytics firm outperformed leading economists in forecasting inflation by blending high-quality, diverse data sources (including real-time consumer trends and commodity prices) into its machine learning models (www.forbes.com [8]). The firm's CEO stressed that the critical breakthrough wasn't a bigger algorithm at all - it was the incorporation of domain expertise and representative, quality data so the AI's predictions would "reflect reality" (www.forbes.com [9]) (www.forbes.com [10]). The lesson is clear: improving data quality, completeness, and governance can determine whether AI initiatives succeed or stall.

References:

[1] www.bain.com — https://www.bain.com/insights/why-ai-stumbles-without-a-solid-data-strategy/#:~:text=a%20Solid%20Data%20Strategy%20At,to%20them%2C%20and%20critical%20to
[2] www.valuebound.com — https://www.valuebound.com/resources/blog/ai-projects-fail-enterprises-2026-reality-check#:~:text=dies%20quietly,Vendors%20promise%20governance%20but
[3] www.forbes.com — https://www.forbes.com/sites/garydrenik/2025/10/15/why-95-of-ai-projects-fail-and-how-better-data-can-change-that/#:~:text=predictive%20systems%20are%20keeping%20vital,structured%2C%20and%20timely%20enough%20with
[4] www.forbes.com — https://www.forbes.com/sites/garydrenik/2025/10/15/why-95-of-ai-projects-fail-and-how-better-data-can-change-that/#:~:text=models%2C%20most%20enterprises%20are%20still,similarly%20cautioned%20that%20data%20provenance
[5] www.forbes.com — https://www.forbes.com/sites/garydrenik/2025/10/15/why-95-of-ai-projects-fail-and-how-better-data-can-change-that/#:~:text=incomplete%2C%20biased%2C%20or%20stale%20datasets%2C,to%20AI%20trustworthiness%20as%20the
[6] www.healthittek.com — https://www.healthittek.com/post/data-governance-or-data-chaos-why-72-of-healthcare-ai-projects-fail-and-how-to-be-inthe-28-that#:~:text=represents%20a%20common%20pattern%20across,This%20data
[7] www.healthittek.com — https://www.healthittek.com/post/data-governance-or-data-chaos-why-72-of-healthcare-ai-projects-fail-and-how-to-be-inthe-28-that#:~:text=information%20effectively,average%20of%20%2415%20million%20annually
[8] www.forbes.com — https://www.forbes.com/sites/garydrenik/2025/10/15/why-95-of-ai-projects-fail-and-how-better-data-can-change-that/#:~:text=the%2095,And%20this
[9] www.forbes.com — https://www.forbes.com/sites/garydrenik/2025/10/15/why-95-of-ai-projects-fail-and-how-better-data-can-change-that/#:~:text=What%20makes%20this%20notable%20is,inputs%2C%20rather%20than%20sheer%20computing
[10] www.forbes.com — https://www.forbes.com/sites/garydrenik/2025/10/15/why-95-of-ai-projects-fail-and-how-better-data-can-

change-that/#:~:text=of%20a%20data,you%20get%20consistent%20accuracy%2C%20not

Proprietary Data as a competitive moat

Facing these data challenges, companies are turning their information into an advantage by treating it as proprietary intellectual property. Leading organisations regard their accumulated data - from customer behaviour and transactions to internal processes - as strategic capital that can fuel AI innovations and differentiate them in the market (www.bain.com [1]). They are investing in “reusable data products” with clear ownership, integrating data from across silos, and implementing rigorous governance so that their AI systems can trust and easily access high-value information from a single source of truth (www.bain.com [2]).

This shift reflects a new competitive reality: data has become the key to sustaining an edge in the AI era. In a recent global survey, 78% of Chief Data Officers said exploiting proprietary data is now a top strategy to outpace competitors (newsroom.ibm.com [3]). As advanced AI models become widely accessible to all, the unique datasets and historical knowledge within an organisation are increasingly seen as the hardest-to-replicate asset - acting as a data moat that protects the business. Companies are beginning to identify which data assets are truly unique and critical for their future, and they are doubling down on protecting and exploiting those crown jewels (www.bain.com [4]).

C-suite data leaders are also addressing the gap between AI ambitions and data readiness. Many report that while data strategy is finally being woven into overall technology roadmaps (81% of organisations, up from just 52% in 2023), only 26% are fully confident that their data is prepared to support new AI-driven revenue streams (newsroom.ibm.com [5]). Nearly half of CDOs also say a lack of advanced data skills in their workforce is now a major challenge (up from 32% two years ago) (newsroom.ibm.com [6]). These realities are reshaping the CTO/CDO agenda, pushing leaders to prioritise data quality initiatives, modern data platform upgrades, and talent development in data management. By doing so, enterprises can turn what was once a data bottleneck into a source of sustained competitive advantage.

References:

[1] www.bain.com — https://www.bain.com/insights/why-ai-stumbles-without-a-solid-data-strategy/#:~:text=images%2C%20and%20video,to%20them%2C%20and%20critical%20to

[2] www.bain.com — https://www.bain.com/insights/why-ai-stumbles-without-a-solid-data-strategy/#:~:text=images%2C%20and%20video,to%20them%2C%20and%20critical%20to

[3] newsroom.ibm.com — https://newsroom.ibm.com/2025-11-13-ibm-study-chief-data-officers-redefine-strategies-as-ai-ambitions-outpace-readiness#:~:text=Outpace%20Readiness%20,of%20respondents%20identify%20advanced%20data

[4] www.bain.com — https://www.bain.com/insights/why-ai-stumbles-without-a-solid-data-strategy/#:~:text=images%2C%20and%20video,to%20them%2C%20and%20critical%20to

[5] newsroom.ibm.com — https://newsroom.ibm.com/2025-11-13-ibm-study-chief-data-officers-redefine-strategies-as-ai-ambitions-outpace-readiness#:~:text=widening%20gap%20between%20AI%20ambition,reach%2C%20but%20success%20depends%20on

[6] newsroom.ibm.com — https://newsroom.ibm.com/2025-11-13-ibm-study-chief-data-officers-redefine-strategies-as-ai-ambitions-outpace-readiness#:~:text=Outpace%20Readiness%20,Y

Next-Gen Data architecture: from lakehouse to real-time AI

The past week brought fresh evidence that AI’s hunger for data is spurring a renaissance in enterprise data architecture. Analysts predict that by the end of 2026, 40% of large-scale applications will embed AI assistants or agents, yet most current data pipelines were never designed for the speed and complexity of these workloads (dataforest.ai [1]). This is accelerating the shift from the old divide of data warehouses versus data lakes to modern “lakehouse” designs - unified platforms that combine the rigorous management and fast queries of traditional warehouses with the flexibility and scale of data lakes, all under one governance framework. By converging previously siloed data stores, lakehouse architectures let companies feed AI models with both real-time streams and historical datasets without cumbersome integrations or copy delays.

At the same time, specialised vector databases have risen to prominence as firms grapple with the 80% of enterprise information that's unstructured "dark data" (text documents, images, emails, call transcripts, etc.) (www.tmcnet.com [2]). These systems store information as numerical embeddings - mathematical vectors that encode semantic meaning - enabling lightning-fast similarity search through vast troves of text and media. This capability has quickly moved from experimental to mission-critical: retrieval-augmented generation (RAG) has become a dominant AI architecture for grounding LLMs in an organisation's own knowledge, and vector databases serve as the core retrieval layer for those systems (www.marktechpost.com [3]).

To meet these needs, vendors are rapidly launching new solutions that unify data storage and access for AI. Last week, Zilliz - the company behind the Milvus open-source vector database used by over 10,000 organisations - announced "Vector Lakebase," a platform that merges a high-performance vector search engine with lakehouse-style shared storage (www.morningstar.com [4]) (www.morningstar.com [5]). The goal is to offer a single "zero-copy" source where the same corpus of data can simultaneously support real-time AI queries, interactive analytics, and even petabyte-scale model training - all without creating extra data copies or pipelines (www.morningstar.com [6]) (www.morningstar.com [7]). Similarly, startups are tackling the latency problem by processing only new or changed data instead of full data reprocessing, enabling millisecond-level updates for AI models and breaking the cost and speed limits of traditional batch pipelines (www.tmcnet.com [8]).

Crucially, many next-gen data platforms embrace open standards so enterprises maintain control and compliance across diverse environments. By leveraging open table formats and semantic metadata layers that capture data lineage and policies, companies ensure their AI systems preserve data sovereignty and transparency by design (www.tmcnet.com [9]). In practice, this means an AI application can trace the origin of the data it uses, respect jurisdictional restrictions on data, and automatically enforce privacy or usage rules - capabilities that reduce risk and build trust in AI-driven operations.

References:

[1] dataforest.ai — https://dataforest.ai/blog/state-of-modern-data-architecture-benchmark-report#:~:text=Research%20Insights%29,in%20practice%2C%20what%20open%20table

[2] www.tmcnet.com — https://www.tmcnet.com/usubmit/2026/06/08/10396030.htm#:~:text=Singdata%27s%20core%20innovations%20%E2%80%94%20Generic,level

[3] www.marktechpost.com — https://www.marktechpost.com/2026/05/10/best-vector-databases-in-2026-pricing-scale-limits-and-architecture-tradeoffs-across-nine-leading-systems/#:~:text=Ever%20in%202026%20The%20shift,that%20LLMs%20query%20at%20inference

[4] www.morningstar.com — https://www.morningstar.com/news/pr-newswire/20260610cn80329/zilliz-launches-vector-lakebase-extending-the-worlds-most-adopted-vector-database-into-a-unified-data-platform-for-ai#:~:text=adopted%20open,Zilliz%20powers

[5] www.morningstar.com — https://www.morningstar.com/news/pr-newswire/20260610cn80329/zilliz-launches-vector-lakebase-extending-the-worlds-most-adopted-vector-database-into-a-unified-data-platform-for-ai#:~:text=every%20release%2C,Why%20a

[6] www.morningstar.com — https://www.morningstar.com/news/pr-newswire/20260610cn80329/zilliz-launches-vector-lakebase-extending-the-worlds-most-adopted-vector-database-into-a-unified-data-platform-for-ai#:~:text=every%20release%2C,Why%20a

[7] www.morningstar.com — https://www.morningstar.com/news/pr-newswire/20260610cn80329/zilliz-launches-vector-lakebase-extending-the-worlds-most-adopted-vector-database-into-a-unified-data-platform-for-ai#:~:text=valuable%20data%20retrievable%20but%20never,asked%20for%20a%20way%20to

[8] www.tmcnet.com — https://www.tmcnet.com/usubmit/2026/06/08/10396030.htm#:~:text=and%20other%20unstructured%20,Pacific

[9] www.tmcnet.com — https://www.tmcnet.com/usubmit/2026/06/08/10396030.htm#:~:text=Iceberg%20open%20standard%2C%20ensuring%20customers,Enterprise%20Validation%20Singdata%27s%20customer%20base

Governance & ethics: no more Data shortcuts

Competitive advantage isn't the only reason to shore up data foundations - regulators and society are also raising the bar for AI. Europe has led the way: the European Union's sweeping AI Act will require any "high-risk" AI system (e.g., in hiring, lending, or healthcare) to meet strict standards for data transparency, quality, and accountability when it takes effect in August 2026. Penalties for non-compliance can reach €35 million or 7% of global annual turnover. However, a recent survey indicates

that 78% of enterprises are not yet prepared to comply with these new AI governance requirements - a clear sign that many firms must urgently strengthen their data practices.

In addition to government rules, companies are facing pressure to prevent privacy breaches and biased AI outcomes. Existing laws like GDPR already impose hefty fines for mishandling personal data, and customers now demand greater transparency into how their information is used by AI systems. High-profile incidents also underscore the need for rigorous internal controls. For instance, after Samsung engineers inadvertently leaked confidential code to a public AI chatbot, the company banned employees from using generative AI tools like ChatGPT altogether (www.cyber-gear.ai [1]). Such episodes highlight the danger of “shadow AI” - unsanctioned use of AI applications - and the importance of strong data governance and employee training to avoid accidental leaks or compliance violations.

Across industries, data sovereignty is becoming a board-level concern that directly influences architecture choices. Nations from China to Saudi Arabia have introduced laws requiring certain data to be stored and processed within their borders. Even in technology hubs like Hong Kong, enterprise platforms now emphasise that customers maintain 100% ownership and control of their own data, reflecting how critical local compliance has become (www.tmcnet.com [2]). For global enterprises, this patchwork of regulations means designing systems that can segregate and monitor data by jurisdiction, and thoroughly document AI training data to ensure accountability.

The bottom line: there are no shortcuts around data if businesses want sustainable AI success. Whether to unlock real economic value or to meet new legal and ethical obligations, investing in data quality, integration, and stewardship has become a strategic imperative. Companies that once saw data as mere “tech plumbing” must now treat it as proprietary capital - a long-term asset that will determine who thrives in the era of AI.

References:

[1] www.cyber-gear.ai — https://www.cyber-gear.ai/samsung-bans-staffs-ai-use-after-spotting-chatgpt-data-leak/#:~:text=Samsung%20Bans%20Staff%E2%80%99s%20AI%20Use,ChatGPT%20after%20discovering%20staff%20uploaded
[2] www.tmcnet.com — https://www.tmcnet.com/usubmit/2026/06/08/10396030.htm#:~:text=Iceberg%20open%20standard%2C%20ensuring%20customers,Enterprise%20Validation%20Singdata%27s%20customer%20base

Key Statistics

- Up to 95% of AI projects fail to deliver on their promises ([www.forbes.com](https://www.forbes.com/sites/garydrenik/2025/10/15/why-95-of-ai-projects-fail-and-how-better-data-can-change-that/#:~:text=predictive%20systems%20are%20keeping%20vital,structured%2C%20and%20timely%20enough%20with)).
- Poor data quality costs businesses about $12.9 million annually, contributing to 40% of failed business initiatives ([www.forbes.com](https://www.forbes.com/sites/garydrenik/2025/10/15/why-95-of-ai-projects-fail-and-how-better-data-can-change-that/#:~:text=incomplete%2C%20biased%2C%20or%20stale%20datasets%2C,to%20AI%20trustworthiness%20as%20the)).
- 78% of surveyed Chief Data Officers say leveraging proprietary data is a top strategic priority for market differentiation ([newsroom.ibm.com](https://newsroom.ibm.com/2025-11-13-ibm-study-chief-data-officers-redefine-strategies-as-ai-ambitions-outpace-readiness#:~:text=Outpace%20Readiness%20,of%20respondents%20identify%20advanced%20data)).
- Only 26% of CDOs are confident their data is ready to support new AI-enabled revenue streams ([newsroom.ibm.com](https://newsroom.ibm.com/2025-11-13-ibm-study-chief-data-officers-redefine-strategies-as-ai-ambitions-outpace-readiness#:~:text=widening%20gap%20between%20AI%20ambition,reach%2C%20but%20success%20depends%20on)).

KEY TAKEAWAY

For C-level leaders, the message is clear: fix your data or risk AI failure. Modernising data architecture, improving data quality and governance, and leveraging unique data assets are now essential to turning AI investments into real ROI - and to meeting growing regulatory demands.

Sources

Why 95% Of AI Projects Fail And How Better Data Can Change That
https://www.forbes.com/sites/garydrenik/2025/10/15/why-95-of-ai-projects-fail-and-how-better-data-can-change-that/

AI Projects Fail in Enterprises: 2026 Reality Check
https://www.valuebound.com/resources/blog/ai-projects-fail-enterprises-2026-reality-check

Why AI Stumbles Without a Solid Data Strategy
https://www.bain.com/insights/why-ai-stumbles-without-a-solid-data-strategy/

IBM Study: Chief Data Officers Redefine Strategies as AI Ambitions Outpace Readiness
https://www.prnewswire.com/news-releases/ibm-study-chief-data-officers-redefine-strategies-as-ai-ambitions-outpace-readiness-302613794.html

The Real Bottleneck in the Age of AI Agents Isn't the Model -- It's the Data
https://www.tmcnet.com/usubmit/2026/06/08/10396030.htm

Zilliz Launches Vector Lakebase, Extending the World's Most Adopted Vector Database into a Unified Data Platform for AI
https://www.prnewswire.com/news-releases/zilliz-launches-vector-lakebase-extending-the-worlds-most-adopted-vector-database-into-a-unified-data-platform-for-ai-302796419.html

New Reports Identify Traits of Enterprise AI Leaders and Laggards
https://redmondmag.com/articles/2025/06/20/new-reports-identify-traits-of-enterprise-ai-leaders-and-laggards.aspx

Cisco research: A major infrastructure shift is underway. AI could double the strain or solve it
https://newsroom.cisco.com/c/r/newsroom/en/us/a/y2025/m06/cisco-research-a-major-infrastructure-shift-is-underway-ai-could-double-the-strain-or-solve-it.html

Best Vector Databases in 2026: Pricing, Scale Limits, and Architecture Tradeoffs Across Nine Leading Systems
https://www.marktechpost.com/2026/05/10/best-vector-databases-in-2026-pricing-scale-limits-and-architecture-tradeoffs-across-nine-leading-systems/

Samsung Bans Staff’s AI Use After Spotting ChatGPT Data Leak
https://www.cyber-gear.ai/samsung-bans-staffs-ai-use-after-spotting-chatgpt-data-leak/