LUMO INSIGHTS

Thursday, 2 July 2026

AI's new moat: Data strategy as the deciding factor

Executive Summary

This week's top developments show that data - not algorithms - is increasingly the make-or-break factor in enterprise AI success. Organisations with strong, well-governed data foundations are pulling ahead, while those that neglect data quality, architecture, and compliance risk falling behind. From eye-opening failure stats to big moves by data-savvy companies, the message is clear: AI readiness depends on getting your data house in order.

AI leaders vs laggards: The Data architecture divide

(campustechnology.com [1])A new industry report confirms that the key difference between organisations successfully scaling AI and those stuck in pilot mode is the strength of their data architecture and governance. TDWI's latest AI-Ready Data Foundation study finds companies seeing the greatest AI impact have invested heavily in integrated data pipelines, unified platforms, and robust governance - far more so than their lower-performing peers (campustechnology.com [2]). In practice, leading AI adopters unify data across silos and enforce consistent definitions and quality standards organisation-wide, ensuring their models train on a single source of truth.

Technology providers are responding to this need for better data foundations. At its annual summit in late June, Snowflake announced a new open framework aimed at eliminating data fragmentation by adopting open table formats like Apache Iceberg and a universal governance catalogue (markets.ft.com [3]). This approach allows teams - and even AI agents - to access a single, live, governed copy of enterprise data wherever it resides without cumbersome duplication (markets.ft.com [4]). Major companies such as Affirm, NTT Docomo, and Samsung Ads are already leveraging Snowflake's unified data architecture to simplify their systems and build AI on a consistent, trusted base (markets.ft.com [5]). Increasingly, business leaders are making data interoperability and strong data foundations top strategic priorities to avoid falling behind in the AI race.

Moreover, as McKinsey experts note, scaling AI demands connecting all types of information - from structured databases to unstructured text and images - into one governed, reusable repository (www.mckinsey.com [6]). Forward-looking CIOs and CDOs recognise that aggregating data from across the enterprise under a common architecture is essential. By doing so, they enable AI models to draw on all relevant knowledge while preserving context, lineage, and control. This data-first mindset is becoming a hallmark of AI leaders - and it is widening the performance gap between them and late adopters.

References:

[1] campustechnology.com — https://campustechnology.com/articles/2026/06/29/report-ai-impact-starts-with-strong-data-foundation.aspx?p=1#:~:text=strategic%20differentiators%2C%20not%20just%20infrastructure,a%20research%20and%20education%20organization

[2] campustechnology.com — https://campustechnology.com/articles/2026/06/29/report-ai-impact-starts-with-strong-data-

foundation.aspx?p=1#:~:text=report%20is%20titled%2C%20,She%20explains%20how
[3] markets.ft.com — https://markets.ft.com/data/announce/detail?dockey=600-202606020906BIZWIRE_USPRX____20260602_BW477507-1#:~:text=Open%20Framework%20for%20Interoperable%20Enterprise,outside%20of%20Snowflake%2C%20while%20minimizing
[4] markets.ft.com — https://markets.ft.com/data/announce/detail?dockey=600-202606020906BIZWIRE_USPRX____20260602_BW477507-1#:~:text=Open%20Framework%20for%20Interoperable%20Enterprise,outside%20of%20Snowflake%2C%20while%20minimizing
[5] markets.ft.com — https://markets.ft.com/data/announce/detail?dockey=600-202606020906BIZWIRE_USPRX____20260602_BW477507-1#:~:text=Polaris%2C%20for%20bi,that%20redefine%20interoperability%20for%20the
[6] www.mckinsey.com — https://www.mckinsey.com/capabilities/mckinsey-technology/our-insights/ai-data-readiness-the-key-to-scaling-impact#:~:text=readiness%3A%20The%20key%20to%20scaling,DOWNLOADS%20Article%20%2812

Data quality & governance - AI’s achilles heel

Despite rising AI investments, many projects are hitting an old obstacle: poor data quality and governance. Analysts warn of an 'AI ROI cliff' - pilots that perform well in controlled tests but then fail to deliver value in production because real-world data is messy (www.dataversity.net [1]) (www.dataversity.net [2]). In sandbox environments, algorithms may excel, but when confronted with years of inconsistent, duplicate-filled, or siloed enterprise data, they often produce unreliable insights. Users lose trust in these flawed AI outputs and revert to manual processes (www.dataversity.net [3]), causing promising AI initiatives to stall.

New statistics reveal how pervasive this challenge remains. Roughly 80% of AI projects still don’t achieve their intended business objectives (www.folio3.ai [4]), and Gartner estimates that 85% of AI project failures stem from poor data quality or lack of relevant data (www.folio3.ai [5]). In fact, barely half of AI initiatives ever progress from pilot to full production deployment (www.folio3.ai [6]). Many are abandoned due to data privacy hurdles, cost overruns, or unclear ROI. Analysts predict that by the end of 2025, at least 30% of generative AI pilots will be dropped after initial trials, with lack of data readiness a primary culprit (www.gartner.com [7]).

These sobering numbers have put data excellence into sharp focus. Surveys show that a majority of companies keen on generative AI have not yet upgraded their data infrastructure to support it - a risky oversight (www.ciodive.com [8]). As one technology leader noted, bad data will inevitably lead to bad models (www.ciodive.com [9]). The most successful organisations are responding by doubling down on data governance and quality: cleaning and integrating datasets, implementing master data management, and clarifying ownership and stewardship. The goal is to ensure that when AI systems move into production, they draw from accurate, up-to-date, well-understood information. Without this solid foundation, even state-of-the-art AI models will struggle to deliver real business value.

References:

[1] www.dataversity.net — https://www.dataversity.net/articles/why-ai-projects-fail-at-scale-the-data-foundation-enterprise-leaders-overlook/#:~:text=the%20demo%20dazzled%20stakeholders%2C%20and,Here%E2%80%99s%20what%20IT%20and%20RevOps
[2] www.dataversity.net — https://www.dataversity.net/articles/why-ai-projects-fail-at-scale-the-data-foundation-enterprise-leaders-overlook/#:~:text=controlled%20conditions%2C%20limited%20variables%2C%20narrowly,predictions%20unreliable%20at%20the%20source
[3] www.dataversity.net — https://www.dataversity.net/articles/why-ai-projects-fail-at-scale-the-data-foundation-enterprise-leaders-overlook/#:~:text=The%20Three%20Failure%20Modes%20That,aren%E2%80%99t%20just%20partial%3B%20they%E2%80%99re%20potentially
[4] www.folio3.ai — https://www.folio3.ai/blog/ai-project-failure-rate-stats#:~:text=Contents%2080,that%20actually%20succeed%20are%20doing
[5] www.folio3.ai — https://www.folio3.ai/blog/ai-project-failure-rate-stats#:~:text=production,What%20Percentage%20of
[6] www.folio3.ai — https://www.folio3.ai/blog/ai-project-failure-rate-stats#:~:text=2025%29%20%E2%80%A2%2042,project%20failures%20trace%20back%20to
[7] www.gartner.com — https://www.gartner.com/en/newsroom/press-releases/2024-07-29-gartner-predicts-30-percent-of-generative-ai-projects-will-be-abandoned-after-proof-of-concept-by-end-of-2025#:~:text=By%20End%20of%202025%20SYDNEY%2C,%E2%80%9CAfter%20last%20year%27s%20hype%2C%20executives
[8] www.ciodive.com — https://www.ciodive.com/news/loreal-ai-data-strategy-earnings/707810/#:~:text=the%20steps%20to%20put%20their,Future%20Tech%20%E2%80%9CAI%20now%20transcends
[9] www.ciodive.com — https://www.ciodive.com/news/loreal-ai-data-strategy-earnings/707810/#:~:text=the%20steps%20to%20put%20their,and%20IT%20leaders%20expect%20data

Proprietary Data as the new competitive moat

In the race to leverage advanced AI, one fact is becoming clear: models are increasingly commoditised, but proprietary data remains uniquely yours. With open-source and commercial AI models readily available, rivals can often access similar algorithms or pre-trained systems. What they can't access is your organisation's unique trove of data - the customer interactions, domain-specific knowledge, and operational insights that only you possess (www.ibm.com [1]). More and more, companies are treating this proprietary data as strategic intellectual property and a true competitive moat in the AI era.

A vivid example comes from Bloomberg's recent foray into generative AI. The financial information giant developed its own large language model, BloombergGPT, trained on decades of proprietary financial data. The model's underlying technology isn't the main source of Bloomberg's advantage - an open-source LLM trained on the same data might perform similarly - but no competitor can match the 40-year archive of curated financial data behind it (www.searchcans.com [2]). In other words, while anyone can download a powerful AI model, nobody can download Bloomberg's data pipeline. This highlights how a rich, well-maintained dataset can translate into smarter, more context-aware AI solutions that competitors without that data cannot easily replicate.

Companies across industries are taking note. Organisations are racing to accumulate and protect valuable datasets that reflect their customers, products, and operations, knowing these will fuel the next wave of AI capabilities. Many are fine-tuning general AI models with their own data or employing methods like retrieval-augmented generation to inject internal knowledge into AI systems. By infusing AI with private, high-quality data - and managing that data with rigorous governance - enterprises can ensure their AI delivers insights and recommendations that are uniquely tailored to their business. In an era when cutting-edge models are accessible to all, a differentiated data foundation may be the last enduring competitive advantage.

References:
[1] www.ibm.com — https://www.ibm.com/think/insights/proprietary-data-gen-ai-competitive-edge#:~:text=Google%2C%20has%20access%20to%20public,tuning.%201
[2] www.searchcans.com — https://www.searchcans.com/blog/new-moat-proprietary-data-pipelines-defensible/#:~:text=structure%20unique%20data%20to%20feed,massive%20infrastructure%20for%20cleaning%2C%20structuring

Emerging Data platforms: lakehouses and vector search

Realising AI's potential requires next-generation data infrastructure built for scale and versatility. One major trend is the rise of the data lakehouse - an architecture that combines the flexibility of data lakes with the reliability of data warehouses. By using open table formats like Apache Iceberg, cloud data platforms now let companies share and access data across diverse systems while maintaining one source of truth and strong governance (www.businesswire.com [1]). This unified approach means teams can run analytics and machine learning on the same platform, eliminating the delays and errors caused by shuffling data between separate silos.

Another breakthrough is the vector database, a technology purpose-built for AI and unstructured content. Unlike traditional relational databases, vector databases store information as high-dimensional numerical embeddings and excel at similarity search - vital for finding relevant text, images, or audio via AI. Enterprise adoption of vector databases has skyrocketed, growing 377% year-over-year as firms deploy them for tasks like customer support chatbots and knowledge retrieval (www.databricks.com [2]). Major vendors are incorporating vector search into their tools; for example, Salesforce's Data Cloud has added a vector store to help businesses index and query previously untapped unstructured data - often around 90% of all enterprise information - to fuel AI-driven insights

(salesforcedevops.net [3]).

These advances in data architecture are more than just technical upgrades - they address real business needs. By breaking down data silos and enabling context-rich, real-time information retrieval, lakehouse platforms and vector search empower a new class of AI applications. Companies can deliver smarter customer experiences (think AI assistants that truly understand a user's history and documents), make split-second operational decisions with live sensor data, and accelerate innovation by mining vast text and image repositories. The lesson for executives is that staying at the forefront of AI requires investing in these data capabilities now. Organisations building modern, flexible data foundations are moving AI projects from pilot to production faster - and gaining insights that leave less-prepared competitors behind.

References:
[1] www.businesswire.com — https://www.businesswire.com/news/home/20240604251194/en#:~:text=ease%20of%20use%2C%20performance%2C%20governance%2C,diverse%20ecosystem%20of%20contributors%20and
[2] www.databricks.com — https://www.databricks.com/blog/state-ai-enterprise-adoption-growth-trends#:~:text=organizations%20are%20prioritizing%20data%20and,The%20data%20spans%20February%201
[3] salesforcedevops.net — https://salesforcedevops.net/index.php/2024/06/06/salesforce-data-cloud-vector-database/#:~:text=Salesforce%2C%20emphasized%20the%20significance%20of,this%20unstructured%20data%2C%20the%20vector

Regulatory and ethical pressures drive Data discipline

No data strategy is complete without addressing the fast-changing regulatory and ethical landscape of AI. Governments worldwide are introducing rules that dictate how organisations manage data for AI. Europe's flagship AI Act, for example, enters its first enforcement phase in August 2026 with strict requirements for transparency and data control. Companies deploying AI will need to document their training data sources, assess and mitigate risks in high-risk applications like hiring or lending, and ensure compliance with privacy laws such as GDPR. Penalties for non-compliance are severe: the EU AI Act allows fines up to €35 million (or 6% of global revenue) for violations (www.recordinglaw.com [1]). Regulators have already begun cracking down - last year, the European Data Protection Board hit Clearview AI with a €30 million fine for scraping personal images to train its facial recognition algorithm without consent (data-privacy-office.eu [2]).

These pressures are elevating data governance and ethics to the C-suite agenda. Leading organisations are proactively implementing comprehensive AI governance frameworks that cover data privacy, security, quality, and bias mitigation. In fact, improving data governance and literacy ranks among the top priorities for nearly 40% of data leaders this year (www.cdotrends.com [3]). By treating responsible data use not just as a compliance task but as a strategic differentiator, enterprises avoid legal pitfalls while building trust with customers and regulators. In the end, companies that embed strong data ethics and governance into their AI initiatives will be better positioned to innovate confidently and sustainably.

References:
[1] www.recordinglaw.com — https://www.recordinglaw.com/world-laws/world-data-privacy-laws/eu-ai-act-and-data-privacy/#:~:text=EU%20AI%20Act%20bans%20social,and%20carries%20EUR%2035M%20fines
[2] data-privacy-office.eu — https://data-privacy-office.eu/fines-for-gdpr-violations-in-ai-systems-and-how-to-avoid-them/#:~:text=without%20fear%20and%20with%20understanding,December%202024
[3] www.cdotrends.com — https://www.cdotrends.com/sites/default/files/whitepapaer_file/cdo_insights_2024_0.pdf#:~:text=quality%20are%20all%20rising%20on,to%20capitalize%20on%20AI%2C%20as

Key Statistics

- Only 7% of companies have fully scaled AI across their organisation, as data readiness remains a primary constraint ([www.mckinsey.com](https://www.mckinsey.com/capabilities/mckinsey-technology/our-insights/ai-data-readiness-the-key-to-scaling-impact#:~:text=views%20from%20McKinsey%20Technology,Sidebar%20A)).
- 85% of AI project failures are attributed to poor data quality or lack of relevant data ([www.folio3.ai](https://www.folio3.ai/blog/ai-project-failure-rate-stats#:~:text=production,What%20Percentage%20of)).
- 42% of enterprises abandoned at least one AI initiative in 2025 - up from 17% in the previous year ([www.folio3.ai](https://www.folio3.ai/blog/ai-project-failure-rate-stats#:~:text=of%20AI%20projects%20fail%20to,cost%20of%20a%20single%20failed)).
- Enterprise use of vector databases for AI grew 377% year-over-year, driven by retrieval and generative AI use cases ([www.databricks.com](https://www.databricks.com/blog/state-ai-enterprise-adoption-growth-trends#:~:text=organizations%20are%20prioritizing%20data%20and,The%20data%20spans%20February%201)).
- The EU AI Act can impose fines up to €35 million (or 6% of global revenue) for non-compliant AI systems ([www.recordinglaw.com](https://www.recordinglaw.com/world-laws/world-data-privacy-laws/eu-ai-act-and-data-privacy/#:~:text=EU%20AI%20Act%20bans%20social,and%20carries%20EUR%2035M%20fines)).

KEY TAKEAWAY

The most advanced AI is worthless without high-quality, well-governed data. Most AI projects fail due to data issues ([www.folio3.ai](https://www.folio3.ai/blog/ai-project-failure-rate-stats#:~:text=production,What%20Percentage%20of)). With new regulations enforcing data transparency, investing in robust, compliant data foundations is now a C-suite imperative.

Sources

TDWI Blueprint Report 2026 – Building an AI-Ready Data Foundation (Campus Technology, June 29, 2026)
https://campustechnology.com/articles/2026/06/29/report-ai-impact-starts-with-strong-data-foundation.aspx

AI Project Failure Rate in 2026: What the Data Shows (Folio3, 2026)
https://www.folio3.ai/blog/ai-project-failure-rate-stats

The New Moat: Data Pipelines Beat AI Models (SearchCans Blog, Dec 28, 2025)
https://www.searchcans.com/blog/new-moat-proprietary-data-pipelines-defensible/

Snowflake Furthers Leadership as the Best Data Foundation for Enterprises (Business Wire, June 4, 2024)
https://www.businesswire.com/news/home/20240604251194/en

State of AI: Enterprise Adoption & Growth Trends (Databricks Blog, Nov 28, 2025)
https://www.databricks.com/blog/state-ai-enterprise-adoption-growth-trends

Salesforce Data Cloud Vector Database: Bringing Generative AI to the Masses (SalesforceDevops.net, June 6, 2024)
https://salesforcedevops.net/index.php/2024/06/06/salesforce-data-cloud-vector-database/

Snowflake Pioneers New Open Framework for Interoperable Enterprise Data and AI (Business Wire, June 2, 2026)
https://www.businesswire.com/news/home/20260602477507/en/

AI data readiness: The key to scaling impact (McKinsey & Company, June 23, 2026)
https://www.mckinsey.com/capabilities/technology/our-insights/ai-data-readiness-the-key-to-scaling-impact

Why AI Projects Fail at Scale: The Data Foundation Enterprise Leaders Overlook (Dataversity, June 29, 2026)
https://www.dataversity.net/articles/why-ai-projects-fail-at-scale-the-data-foundation-enterprise-leaders-overlook/

Gartner Press Release: 30% of Generative AI Projects Will Be Abandoned After PoC by End of 2025 (July 29, 2024)
https://www.gartner.com/en/newsroom/press-releases/2024-07-29-gartner-predicts-30-percent-of-generative-ai-projects-will-be-abandoned-after-proof-of-concept-by-end-of-2025

L'Oreal leans into AI with data in mind (CIO Dive, Feb 16, 2024)

https://www.ciodive.com/news/loreal-ai-data-strategy-earnings/707810/

Proprietary data — your competitive edge in generative AI (IBM AI in Action 2024 report)
https://www.ibm.com/think/insights/proprietary-data-gen-ai-competitive-edge

How the EU AI Act Supplements GDPR in the Protection of Personal Data (INTA, June 13, 2024)
https://www.inta.org/perspectives/features/how-the-eu-ai-act-supplements-gdpr-in-the-protection-of-personal-data/

Fines for GDPR violations in AI systems and how to avoid them (Data Privacy Office, 2025)
https://data-privacy-office.eu/fines-for-gdpr-violations-in-ai-systems-and-how-to-avoid-them/

CDO Insights 2024: Charting a Course to AI Readiness (Wakefield Research/Informatica)
https://www.cdotrends.com/sites/default/files/whitepapaer_file/cdo_insights_2024_0.pdf