LUMO INSIGHTS

Monday, 8 June 2026

The Data decisions that make or break AI success

Executive Summary

New developments this week highlight a widening gap between AI leaders and laggards, and the difference comes down to data. From quality and governance pitfalls to new tech that unifies and secures information, firms are learning that AI readiness is all about the foundations.

The great AI Data divide

A new wave of industry research underscores a stark reality: in the race to generate value from AI, the biggest differentiator isn't algorithms - it's the strength of an organisation's data foundations. AI adoption is now widespread, yet only a small fraction of companies are reaping significant returns. McKinsey reports that nearly two-thirds of enterprises have experimented with advanced "agentic" AI systems, but fewer than 10% have successfully scaled these systems to deliver tangible business impact (www.mckinsey.com [1]). Weak data management is often to blame - eight in ten companies cite data limitations as a roadblock to scaling such AI solutions (www.mckinsey.com [2]).

New findings quantify how top performers manage data differently. Gartner reveals that organisations with successful AI initiatives invest up to four times more (as a percentage of revenue) in core data and analytics foundations - things like data quality, governance, skilled talent, and change management - than those getting poor results from AI (itnewsonline.com [3]). These unglamorous investments are paying off. Future-built "AI leader" firms have been found to achieve roughly double the revenue growth and 40% greater cost reductions from AI compared to laggards (www.bcg.com [4]). In short, the gap between AI winners and losers is increasingly a data gap - companies that treat data as a strategic asset are pulling ahead, while those that neglect their data architecture find their AI efforts stalling.

References:

[1] www.mckinsey.com — https://www.mckinsey.com/capabilities/mckinsey-technology/our-insights/building-the-foundations-for-agentic-ai-at-scale#:~:text=agentic%20AI%20at%20scale%20,a%20collaborative%20effort%20by%20Asin

[2] www.mckinsey.com — https://www.mckinsey.com/capabilities/mckinsey-technology/our-insights/building-the-foundations-for-agentic-ai-at-scale#:~:text=Nearly%20two,is%20a%20core%20element%20of

[3] itnewsonline.com — https://itnewsonline.com/news/Gartner-Organizations-with-Successful-AI-Initiatives-Invest-up-to-Four-Times-More-in-Data-and-Analytics-Foundations/35892#:~:text=Analytics%20Foundations%20IT%20News%20Online,management%2C%20compared%20to%20those%20that

[4] www.bcg.com — https://www.bcg.com/press/30september2025-ai-leaders-outpace-laggards-revenue-growth-cost-savings#:~:text=early%20enjoy%20outsized%20benefits%20across,year%20total%20shareholder

Governance, quality & the trust paradox

Why are so many would-be AI projects floundering? One major culprit is the state of data governance and quality. Many organisations forged ahead with generative AI pilots last year only to realise their data house was not in order. In Informatica's latest global CDO survey, 76% of data leaders confessed their governance frameworks can't keep up with how employees are using AI in the wild (www.informatica.com [1]). This leads to what experts dub a "trust paradox" (venturebeat.com [2]): enterprises have rapidly deployed AI - 69% say they now use generative AI, and nearly half have some form of autonomous "agentic AI" in operation (venturebeat.com [3]) - yet their people and processes aren't fully prepared. Three-quarters of these data leaders say employees need significant upskilling in data and AI literacy to use AI responsibly (www.informatica.com [4]). In the rush to AI, many organisations skipped strengthening data governance and training, leaving a trust gap between AI systems and the humans who use them.

Data quality issues compound this challenge. Over half (57%) of chief data officers report that poor data quality and "data reliability" are now the top barriers keeping AI pilots from reaching full production and scale (www.informatica.com [5]). A recent RAND Corporation study similarly pointed to inadequate data governance, low data quality, and fragmented integration as primary factors behind the high failure rate - estimated at over 80% - of corporate AI initiatives (venturebeat.com [6]). When data is fragmented, biased, or unverified, AI models cannot be trusted, and projects struggle to get off the ground. It's an all-too-common story: ambitious AI prototypes produce tantalizing proofs of concept, but then falter in the real world because the data underlying them is incomplete, inconsistent, or non-compliant.

These gaps have very real consequences for the bottom line. One dire prediction from Gartner holds that through 2026, 60% of AI projects that lack sufficient "AI-ready" data will ultimately be abandoned by organisations (www.gartner.com [7]). Or, as a leading Gartner analyst put it, "Without trust in the data, outputs and decisions of AI models and agents… there is no value from AI." (www.webpronews.com [8]) Put simply, without high-quality, well-governed data, even the most sophisticated AI tools cannot deliver sustained business value.

References:

[1] www.informatica.com — https://www.informatica.com/about-us/news/news-releases/2026/01/20260127-new-global-cdo-report-reveals-data-governance-and-ai-literacy-as-key-accelerators-in-ai-adoption.html#:~:text=three,of%20data%20leaders%20say%20their

[2] venturebeat.com — https://venturebeat.com/data/the-trust-paradox-killing-ai-at-scale-76-of-data-leaders-cant-govern-what#:~:text=dangerous%20disconnect%20that%20explains%20why,trust%20the%20data%20powering%20AI

[3] venturebeat.com — https://venturebeat.com/data/the-trust-paradox-killing-ai-at-scale-76-of-data-leaders-cant-govern-what#:~:text=dangerous%20disconnect%20that%20explains%20why,trust%20the%20data%20powering%20AI

[4] www.informatica.com — https://www.informatica.com/about-us/news/news-releases/2026/01/20260127-new-global-cdo-report-reveals-data-governance-and-ai-literacy-as-key-accelerators-in-ai-adoption.html#:~:text=related%20to%20privacy%2C%20security%20and,of%20data%20leaders%20say%20their

[5] www.informatica.com — https://www.informatica.com/about-us/news/news-releases/2026/01/20260127-new-global-cdo-report-reveals-data-governance-and-ai-literacy-as-key-accelerators-in-ai-adoption.html#:~:text=responsible%20and%20effective%20AI%20use,or%20nearly%20all%20their%20employees

[6] venturebeat.com — https://venturebeat.com/ai/salesforce-launches-ai-trust-layer-to-tackle-enterprise-deployment-failures#:~:text=where%20AI%20agents%20work%20alongside,This

[7] www.gartner.com — https://www.gartner.com/en/newsroom/press-releases/2025-02-26-lack-of-ai-ready-data-puts-ai-projects-at-risk#:~:text=predicts%20that%20through%202026%2C%20organizations,and%20govern%20AI%20across%20the

[8] www.webpronews.com — https://www.webpronews.com/ais-silent-force-quadruple-investments-in-data-core-separate-winners-from-laggards/#:~:text=Sallam%20warns%3A%20%E2%80%9CWithout%20trust%20in,of%20firms%20have%20grown%20their

Data as a competitive moat

The flip side of these cautionary tales is the growing recognition that proprietary data itself is becoming a primary source of competitive advantage. In 2026, many industry strategists argue that proprietary data - even more than superior algorithms - is now the strongest and most durable AI “moat” a company can have (aiireland.ie [1]). The reasoning is clear: state-of-the-art AI models are increasingly available to all, whether via open-source communities or cloud AI services, but a unique trove of high-quality data remains inimitable. Boards that fail to govern, protect, and capitalise on their organisation’s unique data assets are, in effect, ceding ground to those who will (aiireland.ie [2]).

We are already seeing this play out in the market. Take the financial sector: Bloomberg’s 50-billion-parameter “BloombergGPT” model was trained on decades of proprietary financial data, enabling it to excel on specialised finance tasks no general-purpose AI can replicate (beancount.io [3]). As one analysis noted, Bloomberg built this model for “tasks only they can define” (beancount.io [4]) - a feat made possible by the company’s exclusive data corpus. Likewise, many enterprises are jealously guarding their data; tech giants like OpenAI have cited competitive concerns to justify not disclosing the full training data behind their AI models. And across industries, from banking to healthcare, a now-familiar trend has emerged: organisations are banning staff from feeding confidential information into public AI chatbots, fearing a loss of data control (moveo.ai [5]).

To build these data moats, leading firms are ramping up investment in data management and security. In fact, 90% of organisations have expanded their privacy initiatives specifically because of AI’s demands, according to Cisco’s large-scale 2026 Data & Privacy Benchmark study (www.webpronews.com [6]). Nearly 40% of companies now spend over $5 million annually on privacy programmes - almost triple the share from just two years ago (www.webpronews.com [7]). This surge in spending isn’t just about avoiding regulatory penalties; it’s about enabling the free flow of trusted data within organisations. From tighter access controls to data encryption and ethical AI reviews, companies are recognising that robust data governance and security not only protect them from risks but also enable AI to operate as a true competitive differentiator.

References:

[1] aiireland.ie — https://aiireland.ie/2026/03/25/the-new-moat-why-proprietary-data-is-your-only-durable-competitive-advantage-in-ai/#:~:text=replicate%20is%20your%20data,handing%20their%20competitors%20the%20advantage
[2] aiireland.ie — https://aiireland.ie/2026/03/25/the-new-moat-why-proprietary-data-is-your-only-durable-competitive-advantage-in-ai/#:~:text=replicate%20is%20your%20data,handing%20their%20competitors%20the%20advantage
[3] beancount.io — https://beancount.io/bean-labs/research-logs/2026/05/05/bloomberggpt-large-language-model-finance#:~:text=unverifiable%2C%20but%20they%27re%20also%20the,equally%20from%20financial%20pretraining%2C%20and
[4] beancount.io — https://beancount.io/bean-labs/research-logs/2026/05/05/bloomberggpt-large-language-model-finance#:~:text=unverifiable%2C%20but%20they%27re%20also%20the,equally%20from%20financial%20pretraining%2C%20and
[5] moveo.ai — https://moveo.ai/blog/companies-that-banned-chatgpt#:~:text=incidents%20within%20a%2020,side%20bans
[6] www.webpronews.com — https://www.webpronews.com/ais-silent-force-quadruple-investments-in-data-core-separate-winners-from-laggards/#:~:text=trust%20deficit%20sparks%20a%20broader,in
[7] www.webpronews.com — https://www.webpronews.com/ais-silent-force-quadruple-investments-in-data-core-separate-winners-from-laggards/#:~:text=trust%20deficit%20sparks%20a%20broader,in

Next-Gen Data architecture: AI’s new foundation

Addressing these challenges, tech providers rolled out key data infrastructure innovations in recent days that aim to help enterprises become “AI-ready.” A core theme is unifying data across silos. Cloud data platforms are embracing the “lakehouse” architecture - merging the reliability of data warehouses with the flexibility of data lakes - to create a single source of truth for analytics and AI. For example, Snowflake’s latest enterprise lakehouse platform lets companies easily ingest and govern data from disparate systems and formats in one place (www.snowflake.com [1]). And in a notable partnership, Salesforce and Databricks have enabled zero-ETL data sharing between their two ecosystems

(www.salesforce.com [2]). This means Salesforce's customer Data Cloud and Databricks' Lakehouse can function as one unified data pool, without the need for time-consuming extract/transform/load pipelines. Moves like these tackle a top pain point, as 41% of IT leaders say their data today is too siloed or complex to be useful (www.salesforce.com [3]).

Another priority is serving up real-time data for AI. Increasingly, AI-driven services (think customer support agents or supply chain optimisers) require instant access to current, clean data. In response, enterprise data warehouses are being turbocharged for low-latency streaming. Snowflake's new "Interactive Tables" and "Interactive Warehouses" are designed to deliver sub-second analytics on live data with high concurrency, so insights feel instantaneous within governed systems (itdigest.com [4]). The platform is also rolling out native support for near-real-time streaming data, letting organisations act on live events within seconds by seamlessly integrating flows from Kafka and other data streams (itdigest.com [5]). The message is clear: in an AI-driven world, stale data means missed opportunities, so data architecture must support continuous, real-time insight generation.

Even the fundamental database is evolving under AI's influence. One fast-emerging technology is the vector database - designed for storing and retrieving the "embeddings" that generative AI models use to understand text, images, and other unstructured data. But instead of introducing yet another silo, many vendors are integrating vector search into existing data platforms to keep AI data under the same governance umbrella. Oracle, for instance, just announced general availability of GPU-accelerated vector index generation inside its flagship database, massively speeding up how quickly enterprises can index and query their data for AI applications (blogs.oracle.com [6]). And for organisations with strict data residency needs, new solutions like Actian's VectorAI DB bring semantic search and Retrieval-Augmented Generation (RAG) capabilities directly to where sensitive data lives - whether in on-premises servers or edge devices - so that AI can come to the data without compromising compliance (www.actian.com [7]) (www.actian.com [8]). Together, these advances point to a future in which AI is not a bolt-on tool but a native part of the data architecture. The best data leaders are already redesigning systems to ensure that their data - whether big or small, structured or unstructured, fast or slow - is accessible, trustworthy, and primed for intelligent automation.

References:

[1] www.snowflake.com — https://www.snowflake.com/en/news/press-releases/snowflake-delivers-the-enterprise-lakehouse-with-enhanced-open--data-access-and-flexibility-for-agentic-ai/#:~:text=with%20Snowflake%20Postgres%2C%20bringing%20developers%E2%80%99,of%20their%20data%20from%20disparate

[2] www.salesforce.com — https://www.salesforce.com/news/stories/salesforce-databricks-data-ai-news/#:~:text=2023%204%20min%20read%20Salesforce,With%20these

[3] www.salesforce.com — https://www.salesforce.com/news/stories/salesforce-databricks-data-ai-news/#:~:text=adoption%20surges%2C%2080,Databricks%20Lakehouse%3A%20Bring%20Your

[4] itdigest.com — https://itdigest.com/cloud-computing-mobility/big-data/snowflake-delivers-the-enterprise-lakehouse-with-enhanced-open-data-access-and-flexibility-for-agentic-ai/#:~:text=Snowflake%20tackles%20this%20with%20the,live%20data%20within%20the%20governed

[5] itdigest.com — https://itdigest.com/cloud-computing-mobility/big-data/snowflake-delivers-the-enterprise-lakehouse-with-enhanced-open-data-access-and-flexibility-for-agentic-ai/#:~:text=Moreover%2C%20Snowflake%20is%20introducing%20near,historical%20context%20to%20address%20use

[6] blogs.oracle.com — https://blogs.oracle.com/database/oracle-ai-database-nvidia-collaboration-advances-enterprise-ai-at-nvidia-gtc-2026#:~:text=GPU,applications%20simple%20and%20their%20data

[7] www.actian.com — https://www.actian.com/company/press-releases/actian-launches-vectorai-db-with-22x-faster-vector-search-for-production-ai-anywhere-including-the-edge/#:~:text=audio%2C%20video%2C%20and%20documents,that%20need%20AI%20embedded%20directly

[8] www.actian.com — https://www.actian.com/company/press-releases/actian-launches-vectorai-db-with-22x-faster-vector-search-for-production-ai-anywhere-including-the-edge/#:~:text=allows%20developers%20to%20bring%20AI,Healthcare%20organizations%20managing%20sensitive%20patient

Regulation raises the stakes

Amid this technological progress, global regulators are making data governance an even more pressing concern. Europe this week unveiled a sweeping “Tech Sovereignty” initiative aimed at reducing dependence on foreign technology and keeping key data and infrastructure under EU control (digital-strategy.ec.europa.eu [1]). Part of this package is a proposed Cloud and AI Development Act, which would impose new requirements on how companies manage data for AI - including where sensitive data can be stored and how AI models are trained and monitored for compliance. Meanwhile, regulations like GDPR continue to influence AI strategies worldwide: companies must carefully manage personal data used in AI to avoid hefty fines and reputational damage (as shown when Italy’s privacy watchdog temporarily blocked ChatGPT in 2023 over data violations).

These pressures are forcing C-level executives to prioritise ethical data practices as a strategic imperative, not just a legal checkbox. Data sovereignty requirements - such as mandates to keep certain data within national borders - are already affecting 81% of organisations and adding significant cost and complexity to AI initiatives (www.webpronews.com [2]). Yet there’s a flip side: strong governance can be a competitive advantage. In one global survey, 96% of businesses said that robust privacy protections actually speed up AI innovation, and 95% believe such measures increase customer trust in their AI-driven services (www.webpronews.com [3]). The takeaway for leaders is that navigating regulations requires investment in privacy and transparency, but those investments can pay dividends. Companies that build compliance into their data architectures from the start - by tracking data lineage, controlling access, and ensuring fairness - are not only avoiding risk; they’re also earning customer confidence and positioning themselves to capitalise on AI’s promises.

References:

[1] digital-strategy.ec.europa.eu — https://digital-strategy.ec.europa.eu/en/policies/eu-tech-sovereignty#:~:text=Policies%20Strengthening%20Europe%E2%80%99s%20Tech%20Sovereignty,Development%20Act%20the%20EU%20Open

[2] www.webpronews.com — https://www.webpronews.com/ais-silent-force-quadruple-investments-in-data-core-separate-winners-from-laggards/#:~:text=aids%20AI%20speed%20and%20invention%2C,highlights%20AI%E2%80%99s%20infrastructure%20hunger%E2%80%94kilowatts%2C%20cooling

[3] www.webpronews.com — https://www.webpronews.com/ais-silent-force-quadruple-investments-in-data-core-separate-winners-from-laggards/#:~:text=systems%2C%20framing%20governance%20as%20business,One%20post

Key Statistics

- Nearly 80% of enterprises say limited data access across environments is stalling their AI initiatives ([www.cloudera.com](https://www.cloudera.com/about/news-and-blogs/press-releases/2026-04-14-nearly-80-percent-of-enterprises-say-ai-is-held-back-by-data-access-challenges-cloudera-report-finds.html#:~:text=striking%20paradox%3A%20while%2096,highlights%20an%20emerging%20%E2%80%9CAI%20readiness))
- Less than one in five companies (18%) report that their data is fully governed enterprise-wide ([www.cloudera.com](https://www.cloudera.com/about/news-and-blogs/press-releases/2026-04-14-nearly-80-percent-of-enterprises-say-ai-is-held-back-by-data-access-challenges-cloudera-report-finds.html#:~:text=match%20at%20L37%20and%20consistency,of%20their%20data%20is%20governed))
- 76% of data leaders admit their AI governance frameworks can’t keep up with employees’ use of AI technology ([www.informatica.com](https://www.informatica.com/about-us/news/news-releases/2026/01/20260127-new-global-cdo-report-reveals-data-governance-and-ai-literacy-as-key-accelerators-in-ai-adoption.html#:~:text=three,of%20data%20leaders%20say%20their))
- Organisations with successful AI programmes invest up to 4× more in data quality, governance, talent, and change management than those with poor AI outcomes ([www.dqchannels.com](https://

www.dqchannels.com/news/gartner-ai-report-april-2026-reveals-why-most-ai-investments-still-fall-short-11752572#:~:text=foundations%20separate%20success%20from%20failure,AI%20outcomes%20data%20quality%2C%20where))

- Gartner predicts that through 2026, 60% of AI projects lacking “AI-ready” data will be abandoned by organisations ([www.gartner.com](https://www.gartner.com/en/newsroom/press-releases/2025-02-26-lack-of-ai-ready-data-puts-ai-projects-at-risk#:~:text=predicts%20that%20through%202026%2C%20organizations,and%20govern%20AI%20across%20the))

KEY TAKEAWAY

Fresh evidence shows a robust data foundation is the true decider of AI success. To get real ROI from AI, leaders must invest now in data quality, integration, governance, and compliance - or risk watching their AI initiatives stall out.

Sources

Nearly 80% of Enterprises Say AI Is Held Back by Data Access Challenges, New Cloudera Report Finds
https://www.cloudera.com/about/news-and-blogs/press-releases/2026-04-14-nearly-80-percent-of-enterprises-say-ai-is-held-back-by-data-access-challenges-cloudera-report-finds.html

The trust paradox killing AI at scale: 76% of data leaders can't govern what employees already use
https://venturebeat.com/data/the-trust-paradox-killing-ai-at-scale-76-of-data-leaders-cant-govern-what

Gartner: Organizations with Successful AI Initiatives Invest up to Four Times More in Data and Analytics Foundations
https://itnewsonline.com/news/Gartner-Organizations-with-Successful-AI-Initiatives-Invest-up-to-Four-Times-More-in-Data-and-Analytics-Foundations/35892

AI Fuels Surge in Data Privacy Investments and Redefines Governance, Cisco Reports
https://www.prnewswire.com/news-releases/ai-fuels-surge-in-data-privacy-investments-and-redefines-governance-cisco-reports-302670021.html

Snowflake Delivers the Enterprise Lakehouse with Enhanced Open Data Access and Flexibility for Agentic AI
https://www.snowflake.com/en/news/press-releases/snowflake-delivers-the-enterprise-lakehouse-with-enhanced-open--data-access-and-flexibility-for-agentic-ai/

Oracle AI Database + NVIDIA Collaboration Advances Enterprise AI at NVIDIA GTC 2026
https://blogs.oracle.com/database/oracle-ai-database-nvidia-collaboration-advances-enterprise-ai-at-nvidia-gtc-2026

Actian launches VectorAI DB with 22x faster vector search for production AI anywhere including the edge
https://www.actian.com/company/press-releases/actian-launches-vectorai-db-with-22x-faster-vector-search-for-production-ai-anywhere-including-the-edge/

BloombergGPT and the Limits of Domain-Specific LLMs in Finance
https://beancount.io/bean-labs/research-logs/2026/05/05/bloomberggpt-large-language-model-finance