LUMO INSIGHTS

Monday, 20 July 2026

Why AI leaders Are pulling ahead: it's the Data

Executive Summary

A new wave of reports and enterprise moves in the past week underscores that data strategy and architecture are now the deciding factors in AI success. Business leaders are finding that cutting-edge algorithms alone aren’t enough - it’s the quality, accessibility, and governance of their data that determine whether AI initiatives deliver real business value.

The Data infrastructure divide

Beneath the AI gold rush, a stark data infrastructure divide is becoming evident. As organisations race to embed AI, many find that the bottleneck is no longer building better models, but connecting, contextualising, and governing the data those models need (www.efficientlyconnected.com [1]). Almost all large enterprises now report integrating AI into core business processes and having clear data strategies, yet nearly 4 out of 5 admit their AI initiatives are still constrained by limited access to siloed data across environments (www.cloudera.com [2]). This “AI readiness illusion” (www.cloudera.com [3]) - AI adoption outpacing the data foundations needed for real impact - underscores that enthusiasm alone can’t deliver AI’s full value. In fact, a new survey found only 6% of self-identified AI-leading firms consider their data infrastructure fully prepared for AI, leaving the vast majority with significant gaps to close (www.cdata.com [4]).

In this environment, the difference between AI leaders and laggards comes down to data readiness. High performers have invested in modern, cloud-based data platforms and architectures (such as the lakehouse, which bridges data warehouses and data lakes) to break down silos and ensure data is analytics-ready across the enterprise. One study shows 60% of AI-mature organisations have heavily modernised their data infrastructure, whereas 53% of low-maturity firms say legacy data systems are their biggest AI obstacle (www.efficientlyconnected.com [5]). And although virtually all companies agree that real-time data is critical for advanced AI, 20% concede they still lack real-time integration capabilities (www.efficientlyconnected.com [6]) - an obvious target for improvement as on-demand intelligence becomes a competitive necessity.

Leaders are also closing these gaps through strategic deals and partnerships. For example, SAP this month acquired Dremio, a high-performance data lakehouse platform, to help customers instantly merge SAP and non-SAP data for analytics and AI - with no need for costly data migrations or conversions (news.sap.com [7]). Likewise, Cloudera’s new partnership with VAST Data (supported by NVIDIA’s AI hardware) promises a fully integrated 'silicon-to-application' data stack for private AI clouds (www.cloudera.com [8]). By uniting high-speed storage, computing, and governance in one solution, this approach aims to eliminate bottlenecks between data and machine learning, enabling enterprises to scale from isolated pilots to reliable, production-grade AI services.

References:

[1] www.efficientlyconnected.com — https://www.efficientlyconnected.com/new-cdata-study-reveals-only-6-of-ai-leaders-are-truly-infrastructure-ready/

#:~:text=many%20cases%2C%20model%20development%20is,context%2C%20quality%2C%20and%20governance%20are
[2] 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=leaders%2C%20the%20report%20finds%20that,%E2%80%9CEnterprises%20aren%E2%80%99t%20struggling%20to
[3] 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=Cloudera%20Report%20Finds%20Press%20Release,to%20deliver%20real%20business%20impact
[4] www.cdata.com — https://www.cdata.com/company/press/state-of-ai-data-connectivity-report/#:~:text=the%20strongest%20predictor%20of%20AI,new%20report%2C%20The%20State%20of
[5] www.efficientlyconnected.com — https://www.efficientlyconnected.com/new-cdata-study-reveals-only-6-of-ai-leaders-are-truly-infrastructure-ready/#:~:text=%2A%2060%25%20of%20high,time%20capabilities
[6] www.efficientlyconnected.com — https://www.efficientlyconnected.com/new-cdata-study-reveals-only-6-of-ai-leaders-are-truly-infrastructure-ready/#:~:text=primary%20blocker%2C%20and%20%2A%20100,time%20capabilities
[7] news.sap.com — https://news.sap.com/2026/07/sap-completes-dremio-acquisition/#:~:text=Acquisition%20of%20Dremio%20Newsbyte%20by,For
[8] www.cloudera.com — https://www.cloudera.com/about/news-and-blogs/press-releases/2026-07-14-cloudera-and-vast-data-announce-strategic-partnership-to-deliver-ai-data-platform-anywhere.html#:~:text=also%20furthers%20solutions%20that%20create,Engineering%20workloads%20by%20allowing%20Spark

Data quality & governance - the hidden bottleneck

No matter how powerful the algorithm, poor data can derail an AI initiative. Studies in 2026 have found that over 80% of enterprise AI projects fail to meet their objectives - about twice the failure rate of traditional IT projects (www.pertamapartners.com [1]). Gartner warns that 60% of AI projects lacking proper data foundations will be abandoned by the end of 2026, and 42% of U.S. enterprises have already cancelled at least one AI initiative due to data quality or integration issues (www.beri.net [2]). These sobering numbers have data leaders asking not “Which model should we build?”, but rather “Are our data and processes ready for AI?”.

Digging into the causes of failure reveals that data issues - not algorithms - are often to blame. In one field study of AI deployments, only about 6% of failures were attributed to the model itself, whereas 31% were caused by integration breakdowns and 22% by “dirty” or inconsistent data (radixweb.com [3]). Put another way, many AI systems stumble because critical data is fragmented, low quality, or lacking proper governance. As a ServiceNow report bluntly observed, most enterprise AI doesn’t fail due to faulty models at all - it fails “because the data is fragmented across disconnected systems and ungoverned,” leading to “shallow intelligence that recommends rather than executes” (investor.servicenow.com [4]).

To avoid these pitfalls, leading organisations are fortifying their data foundations before scaling up AI. They are implementing enterprise-wide governance frameworks, cleansing and unifying siloed datasets, and establishing clear data ownership to improve quality and trust. The logic is simple: AI projects fed with unreliable data will only generate faster mistakes. Conversely, with high-quality, trusted data and strong governance in place, companies see faster returns and lower risk from AI, whereas without these foundations even advanced solutions will likely fail to deliver value (aijourn.com [5]).

References:
[1] www.pertamapartners.com — https://www.pertamapartners.com/insights/ai-project-failure-statistics-2026#:~:text=%2A%201.More%20than%2080,MIT%20Project%20NANDA%2C%202025
[2] www.beri.net — https://www.beri.net/article/ai-project-failure-complete-guide-2026#:~:text=Gartner%27s%20prediction%3A%2060,project%20due%20to%20data%20issues
[3] radixweb.com — https://radixweb.com/ai-failure-report#:~:text=break%20Failure%20Zone%20Share%20of,caused%20by%20technical%20model%20drift
[4] investor.servicenow.com — https://investor.servicenow.com/news/news-details/2026/ServiceNow-launches-the-real-time-data-foundation-that-puts-autonomous-AI-to-work-across-the-enterprise/default.aspx#:~:text=AI%20requires,Context%20Engine%20and%20Autonomous%20Data
[5] aijourn.com — https://aijourn.com/data-priorities-2026-ai-adoption-exposes-gaps-in-data-quality-governance-and-literacy-says-info-tech-research-group-in-new-report/#:~:text=clarity%2C%20and%20they%E2%80%99ll%20scale%20intelligence%2C%E2%80%9D,Tech%E2%80%99s%20Four

Data as moat, Data as mandate

Forward-looking companies now view their data as both a competitive asset and a source of new obligations. An IBM global study of Chief Data Officers (CDOs) found 84% have gained significant competitive advantage from their organisations' proprietary data products (newsroom.ibm.com [1]). Furthermore, 78% of CDOs said leveraging proprietary data is a top strategic objective to differentiate in the market (newsroom.ibm.com [2]). With cutting-edge AI models becoming widely accessible, unique, high-quality data - from customer interactions to industry-specific knowledge - is increasingly the primary moat separating winners from also-rans.

Market evidence backs this up: by 2026, companies with valuable proprietary AI training datasets have been commanding 3 - 5× higher valuations than peers without those assets (fourweekmba.com [3]). This rising "data moat" effect is pushing businesses to invest heavily in capturing and curating data as intellectual property. At the same time, concerns over data ownership and privacy are mounting. Many firms are wary of sharing their crown-jewel datasets with third-party AI vendors, fearing leaks of intellectual property or regulatory breaches. In response, some are prioritising in-house AI development and "sovereign AI" architectures that keep sensitive data under strict control (www.cloudera.com [4]).

Regulators are also raising the stakes. Europe's far-reaching AI Act, for instance, begins enforcement on August 2, 2026, imposing rigorous requirements for data transparency and governance in high-risk AI systems - with penalties up to €35 /million or 7% of global revenue for non-compliance. And just last week, Europe's Data Protection Board called for a legal framework to enable cross-agency information sharing, aiming to strengthen oversight of AI and data across sectors (www.edpb.europa.eu [5]). The message is that enterprises must double down on data governance, documentation, and sovereignty compliance, even as they leverage data for competitive gain. In the AI-driven economy, data has truly become a double-edged sword - the source of innovation and advantage, but also a heightened source of regulatory and ethical risk.

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=ahead%20of%20governance%20and%20security,AI%20ambitions%20remain

[2] newsroom.ibm.com — https://newsroom.ibm.com/2025-11-13-ibm-study-chief-data-officers-redefine-strategies-as-ai-ambitions-outpace-readiness#:~:text=ahead%20of%20governance%20and%20security,AI%20ambitions%20remain

[3] fourweekmba.com — https://fourweekmba.com/data-moats-the-ultimate-competitive-advantage-in-the-digital-age/#:~:text=Master%20Skill%20,Real

[4] www.cloudera.com — https://www.cloudera.com/about/news-and-blogs/press-releases/2026-07-14-cloudera-and-vast-data-announce-strategic-partnership-to-deliver-ai-data-platform-anywhere.html#:~:text=systems%20that%20continuously%20transform%20enterprise,solutions%2C%20accelerated%20by%20NVIDIA%20NIM

[5] www.edpb.europa.eu — https://www.edpb.europa.eu/news/edpb-calls-for-legal-basis-for-cross-regulatory-information-sharing_en#:~:text=regulatory%20cooperation%20The%20Board%20underlines,regulatory%20cooperation%2C%20including

Key Statistics

- Only 6% of enterprise AI leaders believe their data infrastructure is fully AI-ready ([www.cdata.com](https://www.cdata.com/company/press/state-of-ai-data-connectivity-report/#:~:text=the%20strongest%20predictor%20of%20AI,new%20report%2C%20The%20State%20of))
- Nearly 80% of enterprises say limited data access is constraining 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,is%20only%20as%20effective%20as))
- Over 80% of AI projects fail to deliver their intended business value, roughly twice the failure rate of traditional IT projects ([www.pertamapartners.com](https://www.pertamapartners.com/insights/ai-project-failure-

statistics-2026#:~:text=%2A%201.More%20than%2080,MIT%20Project%20NANDA%2C%202025))

- Less than 20% of organisations report having fully governed enterprise data ([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=accessibility,of%20their%20data%20is%20governed))
- By 2026, companies with proprietary AI training datasets command 3 - 5× higher valuations than competitors ([fourweekmba.com](https://fourweekmba.com/data-moats-the-ultimate-competitive-advantage-in-the-digital-age/#:~:text=Master%20Skill%20,Real))

KEY TAKEAWAY

Data readiness has become the gating factor for AI success. To capture AI's value and avoid failures, organisations must prioritise modern, integrated data infrastructure, high-quality data, and strong governance to drive outcomes and ensure compliance.

Sources

Only 6% of AI Leaders Say Their Data Infrastructure Is Ready for AI
https://www.cdata.com/company/press/state-of-ai-data-connectivity-report/

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

AI Project Failure Statistics 2026: The Complete Picture
https://www.pertamapartners.com/insights/ai-project-failure-statistics-2026

AI Failure Report 2026 | Radixweb's Field Intelligence Study
https://radixweb.com/ai-failure-report

Data Priorities 2026: AI Adoption Exposes Gaps in Data Quality, Governance, and Literacy, Says Info-Tech Research Group in New Report
https://aijourn.com/data-priorities-2026-ai-adoption-exposes-gaps-in-data-quality-governance-and-literacy-says-info-tech-research-group-in-new-report/

SAP Completes Acquisition of Dremio
https://news.sap.com/2026/07/sap-completes-dremio-acquisition/

Cloudera and VAST Data Announce Strategic Partnership to Deliver AI Data Platform Anywhere
https://www.vastdata.com/press-releases/cloudera-and-vast-data-strategic-partnership-to-deliver-ai-data-platform-anywhere

AI Database Revolution: RegattaDB Redefines Vector Database
https://thesiliconreview.com/2026/07/ai-database-regattadb-vector-database-ai-agents

ServiceNow launches the real-time data foundation that puts autonomous AI to work across the enterprise
https://newsroom.servicenow.com/press-releases/details/2026/ServiceNow-launches-the-real-time-data-foundation-that-puts-autonomous-AI-to-work-across-the-enterprise/default.aspx

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

EDPB calls for legal basis for cross-regulatory information sharing (17 July 2026)
https://www.edpb.europa.eu/news/edpb-calls-for-legal-basis-for-cross-regulatory-information-sharing_en

EU AI Act Deadlines 2026-2027: Compliance Calendar + Fines
https://www.legiscope.com/blog/eu-ai-act-timeline-deadlines.html