LUMO INSIGHTS

Thursday, 7 May 2026

Data strategy or bust: how smart Data decisions drive AI success

Executive Summary

A flurry of developments this week drives home that robust data foundations - not just advanced algorithms - are determining which companies achieve breakthrough AI results. Shocking new statistics about AI project failures and big tech moves to eliminate data silos are highlighting the gap between AI ambition and reality. It’s clear that data strategy has become the linchpin of AI readiness and a key competitive advantage.

The Data divide: AI leaders vs. laggards

A wave of new industry research is underscoring a stark reality in the AI race: the top performers are pulling ahead not because they have better algorithms, but because they have better data foundations. Gartner’s latest analysis finds that companies successfully scaling AI spend up to four times more (as a percentage of revenue) on core data quality, governance, and infrastructure than those still struggling (www.webpronews.com [1]). This data-centric investment gap translates directly into results - organisations with mature data setups see up to 65% greater improvements in revenue growth and cost reductions from their AI initiatives compared to peers without strong data foundations (www.webpronews.com [2]).

These findings echo other recent studies indicating that only a small fraction of firms (around 5%) are realising significant AI ROI, while the majority remain mired in pilot projects that aren’t scaling (www.forbes.com [3]). The key differentiator is that AI leaders treat data as a strategic asset. They break down silos, ensure data is clean and well-governed, and integrate their AI solutions deeply into business processes to drive value. Laggards, by contrast, often chase the latest AI tools or models without first solidifying their data foundation - a mistake that leads to underperforming models and disappointing results.

In short, having the most advanced model means little if it’s not fuelled by high-quality, well-integrated data. Enterprises that prioritise robust data architecture and data preparation are finding that AI can truly transform their operations and outcomes. Those that neglect these essentials risk falling behind, regardless of how impressive their algorithms might be.

References:

[1] www.webpronews.com — https://www.webpronews.com/ais-silent-force-quadruple-investments-in-data-core-separate-winners-from-laggards/#:~:text=Companies%20chasing%20artificial%20intelligence%20breakthroughs,performers%20don%E2%80%99t%20just%20throw%20money

[2] www.webpronews.com — https://www.webpronews.com/ais-silent-force-quadruple-investments-in-data-core-separate-winners-from-laggards/#:~:text=technology%20executives%20feel%20sure%20their,Data%20quality%20ensures%20inputs%20are

[3] www.forbes.com — https://www.forbes.com/sites/sylvainduranton/2025/11/20/the-widening-ai-value-gap-leaders-profit-laggards-get-left-behind/#:~:text=a%20new%20report%2C%20The%20Widening,greater%20cost

Data quality & governance - The hidden Bottleneck

Despite the excitement around generative AI, data issues are emerging as a primary bottleneck to success. A recent S&P Global survey of over 1,000 IT and business leaders revealed a sharp rise in AI project attrition: 42% of companies have now abandoned most of their AI initiatives before they ever hit production, up from just 17% the year prior (www.ciodive.com [1]). On average, these organisations report scrapping 46% of their AI proof-of-concepts as they struggle to move from pilot to scale. Experts have been frank about the reason - an MIT study found up to 95% of AI projects fail to meet expectations, and the chief culprit isn't the algorithms at all. It's the data: models too often are fed with poor-quality, incomplete or irrelevant information and thus never deliver useful results (www.forbes.com [2]).

These data shortcomings take many forms. Corporate data is frequently fragmented across silos or locked in legacy systems, depriving AI of the consistent context it needs to generate accurate insights (www.prnewswire.com [3]). Key data may be incomplete, inconsistent, or simply inaccurate, undermining model predictions and eroding trust (newsroom.ibm.com [4]). In many firms, data accessibility and integration remain challenges - data scientists still report spending the bulk of their time cleaning and organising data instead of building AI solutions. All of this means that even the most powerful AI models will produce biased or unhelpful outputs if the input data is a mess.

Recognising this, industry leaders are urging a refocus on data fundamentals. An influential industry round-up this week put it bluntly: "The real challenge of enterprise AI is no longer the capabilities of the models themselves but the readiness of the organisation to manage, govern, and integrate data effectively" (cloudtweaks.com [5]). In fact, analysts predict some high-profile AI initiatives could falter in 2026, prompting a pause for reflection and new investments in "AI-specific" data governance to prevent disasters before they happen (cloudtweaks.com [6]). In short, robust data governance and quality management have moved from a back-office IT concern to a frontline factor deciding AI success.

References:

[1] www.ciodive.com — https://www.ciodive.com/news/AI-project-fail-data-SPGlobal/742590/#:~:text=Intelligence,So

[2] 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=Getty%20Images%20With%20government%20data,So%20Many%20AI%20Projects%20Fail

[3] www.prnewswire.com — https://www.prnewswire.com/news-releases/sap-to-acquire-dremio-to-unify-sap-and-non-sap-data-to-power-agentic-ai-302761280.html#:~:text=pending%20regulatory%20approval,SAP

[4] newsroom.ibm.com — https://newsroom.ibm.com/2025-11-13-ibm-study-chief-data-officers-redefine-strategies-as-ai-ambitions-outpace-readiness#:~:text=roadmap%20and%20infrastructure%20investments%20,just%20improve%20their%20AI%2C%20they%27ll

[5] cloudtweaks.com — https://cloudtweaks.com/2025/11/data-governance-becomes-the-ai-throttle/#:~:text=data%20capabilities%20required%20for%20sustainable,Specific%20Data%20Governance%20Is%20the

[6] cloudtweaks.com — https://cloudtweaks.com/2025/11/data-governance-becomes-the-ai-throttle/#:~:text=Bottleneck%20AI%20is%20going%20to,is%20not%20a%20sign%20of

Proprietary Data as a competitive moat

Amid these cautionary tales, companies with strong data foundations are demonstrating how proprietary data can become a powerful competitive moat in the AI era. Just this week, Thomson Reuters reported a 10% jump in first-quarter revenue, attributing the gain in large part to successful AI-driven products built on the firm's vast troves of legal and financial data (www.alphapilot.tech [1]). Earlier in the year, the company's stock price soared nearly 14% in a single day after it announced a new AI-powered CoCounsel legal assistant alongside a partnership with AI pioneer Anthropic (www.financialcontent.com [2]). Investors saw this as proof that Thomson Reuters' unique data assets and specialised AI tools are not just resilient in a turbulent market - they are fuelling the company's growth and outpacing competitors (www.financialcontent.com [3]).

Notably, the nature of these "data moats" is evolving. As one industry expert observed, the advantage of proprietary data is shifting from sheer volume to a rich context layer built on metadata, governance, and domain-specific knowledge. In practice, this means leading organisations are investing in capabilities like detailed data catalogues, lineage tracking, domain ontologies and knowledge graphs, and even vector databases that enable retrieval-augmented generation (RAG) workflows. By layering context and oversight on top of their raw data, companies make their information far more usable for AI systems - which in turn allows AI to deliver deeper, real-time insights and more trustworthy decisions.

Concerns about data ownership and security are also shaping AI strategies. In sectors like finance, healthcare and defence, a growing number of firms now choose to fine-tune AI models on their own infrastructure or within virtual private clouds, to ensure sensitive customer information and intellectual property never leave their control (bytexel.org [4]). This trend reflects a broader realisation: in the age of generative AI, a company's data is not just an operational resource, but a core piece of intellectual property. Keeping that data both protected and primed for AI use is becoming a strategic imperative.

References:

[1] www.alphapilot.tech — https://www.alphapilot.tech/discover/thomson-reuters-q1-2026-revenue-surges-10-as-ai-integration-drives-organic-growth#:~:text=Thomson%20Reuters%20Corporation%20reported%20a,and%20successful%20AI%20product%20adoption
[2] www.financialcontent.com — https://www.financialcontent.com/article/marketminute-2026-2-24-data-as-the-ultimate-moat-thomson-reuters-surges-14-as-ai-strategy-defies-sector-sell-off#:~:text=Thomson%20Reuters%20,seat%22%20subscription
[3] www.financialcontent.com — https://www.financialcontent.com/article/marketminute-2026-2-24-data-as-the-ultimate-moat-thomson-reuters-surges-14-as-ai-strategy-defies-sector-sell-off#:~:text=While%20the%20market%20has%20grown,Off%20Spell%3A%20The%20Road%20to
[4] bytexel.org — https://bytexel.org/the-fine-tuning-paradox-engineering-proprietary-intelligence-in-2026/#:~:text=2026%20bytexel,never%20leaves%20their%20controlled%20environments

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

The past week also saw major moves to modernise data architecture for AI. A prime example is SAP's announcement of its acquisition of Dremio, an open lakehouse platform, which will help customers unify SAP and non-SAP data on a single platform using the open-source Apache Iceberg format (www.prnewswire.com [1]). By integrating Dremio into SAP's Business Data Cloud, the enterprise software giant aims to eliminate data fragmentation and give AI "agentic" systems high-performance access to governed, real-time data across the business (www.prnewswire.com [2]). As SAP's CTO put it, enterprise AI projects don't stall for lack of sophisticated models, but because "the data isn't ready for AI agents" (www.prnewswire.com [3]) - this deal is about removing that obstacle.

Across the industry, data platform providers are similarly doubling down on "AI-first" architecture. Modern lakehouse designs, popularised by cloud data companies like Databricks and Snowflake, blend the flexibility of data lakes with the governance and speed of data warehouses - a formula now being adopted even by legacy vendors like SAP (www.prnewswire.com [4]). Major cloud platforms are rolling out new capabilities to feed AI models with up-to-date information, moving from batch-processed data to continuous data pipelines and streaming analytics so that AI systems are never out of sync with the business's latest context (cloud.google.com [5]). In addition, the rise of vector databases (purpose-built for storing and retrieving the embeddings that power AI knowledge queries) is giving enterprises a way to provide "memory" to generative AI applications. One sign of this trend: in March, an open-source vector search firm secured $50 million in funding to scale its technology for enterprise AI deployments (tech.eu [6]). It's clear that tools like these - once seen as niche - are quickly becoming core parts of the AI infrastructure stack.

Common to all these innovations is an emphasis on openness and interoperability. Companies are increasingly favouring open standards and cloud-agnostic data architectures to avoid being locked into any one vendor's ecosystem (www.prnewswire.com [7]). By building flexible data foundations (for example, adopting open table formats like Iceberg and making data accessible across multiple cloud environments), enterprises can ensure their AI systems remain adaptable. This future-proofing means that as new AI models and techniques emerge, they can be integrated with the organisation's entire corpus of trusted data, without having to re-engineer or migrate the underlying data layer.

References:
[1] www.prnewswire.com — https://www.prnewswire.com/news-releases/sap-to-acquire-dremio-to-unify-sap-and-non-sap-data-to-power-agentic-ai-302761280.html#:~:text=SAP%20to%20Acquire%20Dremio%20to,projects%20fail%20to%20deliver%20value
[2] www.prnewswire.com — https://www.prnewswire.com/news-releases/sap-to-acquire-dremio-to-unify-sap-and-non-sap-data-to-power-agentic-ai-302761280.html#:~:text=enough%3B%20it%20stalls%20because%20the,agentic%20AI%20at%20enterprise%20scale
[3] www.prnewswire.com — https://www.prnewswire.com/news-releases/sap-to-acquire-dremio-to-unify-sap-and-non-sap-data-to-power-agentic-ai-302761280.html#:~:text=data%20with%20high%20performance%20and,agentic%20AI%20at%20enterprise%20scale
[4] www.prnewswire.com — https://www.prnewswire.com/news-releases/sap-to-acquire-dremio-to-unify-sap-and-non-sap-data-to-power-agentic-ai-302761280.html#:~:text=pending%20regulatory%20approval,SAP
[5] cloud.google.com — https://cloud.google.com/blog/products/data-analytics/the-future-of-data-lakehouse-for-the-agentic-era#:~:text=engineered%20for%20the%20era%20of,performance%2C%20scale%2C%20governance%2C%20and%20multimodal
[6] tech.eu — https://tech.eu/2026/03/12/qdrant-closes-50m-series-b-to-expand-vector-search-infrastructure/#:~:text=Qdrant%2C%20an%20open,Ventures%2C%20Spark%20Capital%2C%20and%2042CAP
[7] www.prnewswire.com — https://www.prnewswire.com/news-releases/sap-to-acquire-dremio-to-unify-sap-and-non-sap-data-to-power-agentic-ai-302761280.html#:~:text=enough%3B%20it%20stalls%20because%20the,agentic%20AI%20at%20enterprise%20scale

Data governance and compliance are now non-negotiable

All these developments are unfolding under intensifying regulatory scrutiny. Data protection authorities are making it clear that trustworthy AI requires strong data governance at every step. In Europe, regulators have already fined companies like Clearview AI for illegally scraping personal images to train face-recognition algorithms (www.autoriteitpersoonsgegevens.nl [1]). Italy's watchdog recently imposed a €5 million penalty on the American company behind the Replika chatbot for mishandling users' personal data in training an AI system (www.edpb.europa.eu [2]). These actions send a message: companies must embed privacy-by-design into their AI data pipelines, or risk severe consequences.

Meanwhile, governments are moving to harden data accountability through new laws. The European Union's AI Act - along with its Digital Services Act, Data Act, and other regulations - is pushing organisations to adopt "governance-by-design," building transparency, fairness, and security into data processes from the ground up. In 2026, with the EU's AI regulatory framework coming into force, the cost of having disorganised or biased "messy data" is no longer just an efficiency issue - it's a legal liability (towardsdatascience.com [3]). Firms will need rigorous documentation of data provenance, quality metrics, and human oversight for high-risk AI systems, or they could face fines and reputational damage.

Another rising priority for CDOs and CTOs is ensuring data and AI sovereignty. Leaders across the globe are recognising that controlling their data - and the infrastructure their AI runs on - is now a strategic necessity, not an afterthought. In one global survey, only 11% to 27% of executives (depending on country) currently see sovereignty over AI and data as "mission-critical," but within three years a majority in major economies are expected to hold that view (venturebeat.com [4]). Notably, the number-one driver behind this push (cited at twice the rate of any other factor) is the need to break data out of silos and maintain direct control, rather than geopolitical concerns (venturebeat.com [5]). In practice, this means more organisations are considering localised cloud

infrastructure, strict data residency controls, and open-source AI models that can be run on in-house data - all in an effort to balance AI innovation with compliance, security, and strategic control of data.

References:

[1] www.autoriteitpersoonsgegevens.nl — https://www.autoriteitpersoonsgegevens.nl/en/documents/decision-fine-clearview-ai#:~:text=Internet%20The%20Autoriteit%20Persoonsgegevens%20,Clearview%20is
[2] www.edpb.europa.eu — https://www.edpb.europa.eu/news/national-news/2025/ai-italian-supervisory-authority-fines-company-behind-chatbot-replika_en#:~:text=,1%20of%20the%20GDPR
[3] towardsdatascience.com — https://towardsdatascience.com/the-2026-data-mandate-is-your-governance-architecture-a-fortress-or-a-liability/#:~:text=and%20the%20Data%20Act%2C%20the,tax%20to%20a%20legal%20liability
[4] venturebeat.com — https://venturebeat.com/ai/ai-and-data-sovereignty-are-now-non-negotiable-for-enterprise-leaders-global-survey-finds/#:~:text=and%20Asia%20Pacific.%20Between%2011,convergence%20of%20data%2C%20AI%2C%20and
[5] venturebeat.com — https://venturebeat.com/ai/ai-and-data-sovereignty-are-now-non-negotiable-for-enterprise-leaders-global-survey-finds/#:~:text=or%20let%20competitors%20define%20the,source%20solutions.%20Third%20most%20often

Key Statistics

- Successful AI firms invest 4× more in data quality and governance than their peers, yielding up to 65% better revenue gains and cost reductions ([www.webpronews.com](https://www.webpronews.com/ais-silent-force-quadruple-investments-in-data-core-separate-winners-from-laggards/#:~:text=Companies%20chasing%20artificial%20intelligence%20breakthroughs,performers%20don%E2%80%99t%20just%20throw%20money)) ([www.webpronews.com](https://www.webpronews.com/ais-silent-force-quadruple-investments-in-data-core-separate-winners-from-laggards/#:~:text=technology%20executives%20feel%20sure%20their,Data%20quality%20ensures%20inputs%20are)).
- 42% of companies abandoned most of their AI projects before deployment, up from 17% the year before, as nearly 46% of AI pilots fail to scale into production ([www.ciodive.com](https://www.ciodive.com/news/AI-project-fail-data-SPGlobal/742590/#:~:text=Intelligence,So)).
- Up to 95% of AI projects fail to meet expectations - and poor data quality is often the primary reason ([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=Getty%20Images%20With%20government%20data,So%20Many%20AI%20Projects%20Fail)).
- 76% of data leaders say their organisation's AI governance has not kept pace with AI adoption ([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)), and 86% plan to increase data management investment in 2026 to improve privacy, governance and data literacy ([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=employees%20need%20upskilling%20in%20data,drivers%20for%20this%20spending%20are)).

KEY TAKEAWAY

The past week shows that business leaders must treat data as a strategic asset for AI. Without major new investments in data quality, governance, integration and security, even the best AI tools will underdeliver - or fail outright.

Sources

AI's Silent Force: Quadruple Investments in Data Core Separate Winners from Laggards - WebProNews
https://www.webpronews.com/ais-silent-force-quadruple-investments-in-data-core-separate-winners-from-laggards/
Why 95% Of AI Projects Fail And How Better Data Can Change That - Forbes
https://www.forbes.com/sites/garydrenik/2025/10/15/why-95-of-ai-projects-fail-and-how-better-data-can-change-that/

AI project failure rates are on the rise: report - CIO Dive
https://www.ciodive.com/news/AI-project-fail-data-SPGlobal/742590/

New Global CDO Report Reveals Data Governance and AI Literacy as Key Accelerators in AI Adoption - Informatica (News Release)
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

Data Governance Becomes the AI Throttle: Why 2026 Will Trigger a Pullback Before Acceleration - CloudTweaks
https://cloudtweaks.com/2025/11/data-governance-becomes-the-ai-throttle/

SAP to Acquire Dremio to Unify SAP and Non-SAP Data to Power Agentic AI - PR Newswire
https://www.prnewswire.com/news-releases/sap-to-acquire-dremio-to-unify-sap-and-non-sap-data-to-power-agentic-ai-302761280.html

The Widening AI Value Gap: Leaders Profit, Laggards Get Left Behind - Forbes
https://www.forbes.com/sites/sylvainduranton/2025/11/20/the-widening-ai-value-gap-leaders-profit-laggards-get-left-behind/

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

The Fine-Tuning Paradox: Engineering Proprietary Intelligence in 2026 - Bytexel Blog
https://bytexel.org/the-fine-tuning-paradox-engineering-proprietary-intelligence-in-2026/

Thomson Reuters Q1 2026 Revenue Surges 10% as AI Integration Drives Organic Growth - AlphaPilot
https://www.alphapilot.tech/discover/thomson-reuters-q1-2026-revenue-surges-10-as-ai-integration-drives-organic-growth

Data as the Ultimate Moat: Thomson Reuters Surges 14% as AI Strategy Defies Sector Sell-Off - MarketMinute (FinancialContent)
https://markets.financialcontent.com/stocks/article/marketminute-2026-2-24-data-as-the-ultimate-moat-thomson-reuters-surges-14-as-ai-strategy-defies-sector-sell-off

Replika's €5 Million GDPR Fine: Key Takeaways for AI Developers - CaptainCompliance
https://captaincompliance.com/education/replikas-e5-million-gdpr-fine-key-takeaways-for-ai-developers/

The 2026 Data Mandate: Is Your Governance Architecture a Fortress or a Liability? - Towards Data Science
https://towardsdatascience.com/the-2026-data-mandate-is-your-governance-architecture-a-fortress-or-a-liability/

AI and data sovereignty are now non-negotiable for enterprise leaders, global survey finds - VentureBeat
https://venturebeat.com/ai/ai-and-data-sovereignty-are-now-non-negotiable-for-enterprise-leaders-global-survey-finds/

Qdrant closes $50M Series B to expand vector search infrastructure - Tech.eu
https://tech.eu/2026/03/12/qdrant-closes-50m-series-b-to-expand-vector-search-infrastructure/

Dutch DPA Fines Clearview AI €30.5M for Illegal Personal Data Collection - Autoriteit Persoonsgegevens (Netherlands)
https://www.autoriteitpersoonsgegevens.nl/en/documents/decision-fine-clearview-ai

AI: Italian Supervisory Authority fines company behind chatbot Replika €5M - EDPB (European Data Protection Board)
https://www.edpb.europa.eu/news/national-news/2025/ai-italian-supervisory-authority-fines-company-behind-chatbot-replika_en