LUMO INSIGHTS

Thursday, 21 May 2026

Data strategy: The decisive factor in AI leadership

Executive Summary

The past week’s developments have underscored a core truth: the biggest barrier to AI success isn’t better algorithms - it’s better data. New research reveals that a vast majority of AI projects are failing due to data quality and governance issues, prompting industry leaders to rethink their data architectures and policies. From major acquisitions aimed at unifying data silos to a growing emphasis on real-time data pipelines, proprietary datasets, and compliance with regulations, companies at the forefront of AI are those making bold moves to get their data foundations right.

The Data trust gap: why AI projects fail

New evidence is shining a light on why so many enterprise AI initiatives under-deliver, and it points straight to data issues. A fresh industry survey found that only about one in five AI projects actually meets its key performance indicators, while the other four out of five are paused, re-tooled, or scrapped entirely (securityboulevard.com [1]). Crucially, the core problems causing these failures are not the AI models themselves. Security leaders who dug into unsuccessful deployments kept uncovering the same culprit: flawed data. In the words of one CISO, when these projects “opened the hood” on failed AI efforts, “it wasn’t the model. It was the data (securityboulevard.com [2]).”

Why is data breaking AI projects? Common issues include "data debt" accumulated over years - data that is unclean, unclassified, or siloed away from those who need it (securityboulevard.com [3]). These issues were often survivable in a pre-AI analytics world, but they become project-killers when trying to scale AI. For example, one company recently launched 11 AI initiatives in a year; only 2 met their goals, and 9 were halted or quietly abandoned. Every one of those failed projects was plagued by poor data quality and known governance gaps (securityboulevard.com [4]). Missing data definitions, duplicate and outdated records, and lax access controls meant that models were being fed unreliable, context-free information - resulting in insights the business could not trust.

These data shortcomings often remain hidden until it’s too late. Teams may blame a misfiring algorithm or a "hallucinating" large language model for an AI project’s poor performance, when the real issue lies in unseen data flaws (securityboulevard.com [5]). Few organisations actively measure data trust or readiness; instead, they track AI activity (like the number of models, queries, or lines of AI-generated code) rather than data quality metrics (securityboulevard.com [6]). This misalignment means that when key outcomes are missed, root causes in the data layer go unnoticed. The result is a costly cycle: AI systems that appear busy in dashboards but fail to deliver meaningful business value.

Smart leaders are learning from these hard truths. Organisations now treating data quality and governance as non-negotiable prerequisites for AI are seeing more projects translate into real results. They start by rigorously assessing and improving the reliability and completeness of data before a single algorithm is deployed (securityboulevard.com [7]). By clearly defining data standards for each AI use case, cleaning and labelling information, and tightening data access and lineage controls, these companies catch problems early - before the AI outputs go off course. The payoff is an AI programme where issues are identified and corrected in pilot phases rather than after full rollout, preventing wasted investment and building greater confidence in AI-driven decisions.

References:

[1] securityboulevard.com — https://securityboulevard.com/2026/05/data-trust-ai-success-blog-series-why-most-ai-projects-are-failing/#:~:text=from%20MIND%E2%80%99s%20research%2C%20,KPIs%3F%20The%20same%20root%20cause

[2] securityboulevard.com — https://securityboulevard.com/2026/05/data-trust-ai-success-blog-series-why-most-ai-projects-are-failing/#:~:text=CISOs%20in%20the%20research%20opened,KPIs%3F%20The%20same%20root%20cause

[3] securityboulevard.com — https://securityboulevard.com/2026/05/data-trust-ai-success-blog-series-why-most-ai-projects-are-failing/#:~:text=matters%3A%20AI%20projects%20rarely%20fail,The%20survey%20data%20confirms%20what

[4] securityboulevard.com — https://securityboulevard.com/2026/05/data-trust-ai-success-blog-series-why-most-ai-projects-are-failing/#:~:text=One%20company%20in%20the%20research,use%20case%2C%20connect%20the%20model

[5] securityboulevard.com — https://securityboulevard.com/2026/05/data-trust-ai-success-blog-series-why-most-ai-projects-are-failing/#:~:text=hidden%20is%20the%20cause,It%E2%80%99s%20also%20how%20a

[6] securityboulevard.com — https://securityboulevard.com/2026/05/data-trust-ai-success-blog-series-why-most-ai-projects-are-failing/#:~:text=hidden%20is%20the%20cause,It%E2%80%99s%20also%20how%20a

[7] securityboulevard.com — https://securityboulevard.com/2026/05/data-trust-ai-success-blog-series-why-most-ai-projects-are-failing/#:~:text=treat%20data%20trust%20as%20a,from%20here%3F%20This%20is%20one

From silos to Lakehouses: unifying Data for AI

A striking theme in enterprise technology news this month is the push to break down data silos as a means to unlock AI at scale. Many companies large and small are grappling with fragmented data landscapes - legacy systems and disparate databases that don’t talk to each other (www.ey.com [1]). The consequence of this fragmentation is that data often remains “locked” in functional or departmental silos and in incompatible formats, stripped of the business context that AI needs to derive insight (news.sap.com [2]). This lack of integration hampers AI initiatives: models trained on narrow, isolated datasets struggle to perform in real-world scenarios, and promising pilot projects can’t scale due to slow data integration and compliance concerns (news.sap.com [3]). Simply put, an AI is only as smart as the data it can readily access.

To address this, leading organisations are overhauling their data architectures with an emphasis on openness and unification. One approach gaining significant traction is the "data lakehouse" - a modern data architecture that combines the rigorous data management of warehouses with the flexibility of data lakes. A recent market study by Forrester found that data lakehouses have become the default architecture for modern analytics and AI workloads (www.dremio.com [4]). The reason is straightforward: they simplify data architectures, reduce costly duplication, and allow enterprises to tap into both structured and unstructured data in one environment. By eliminating unnecessary data movement, teams can train and deploy AI models directly on comprehensive, up-to-date datasets, improving both speed and accuracy of insights.

Major tech investments in the last fortnight reflect this shift towards unified data foundations. For example, enterprise software giant SAP announced a deal to acquire data platform Dremio, aiming to merge data across its SAP Business Data Cloud and any external sources seamlessly (news.sap.com [5]) (news.sap.com [6]). Dremio is a leading proponent of open data formats (notably Apache Iceberg) and "enterprise lakehouse" technology, which will let SAP customers query and analyse SAP and non-SAP data together without cumbersome extract-transfer-load processes or data duplication (news.sap.com [7]). SAP itself acknowledged that many AI projects fail not due to AI,

but because underlying data is fragmented and locked away in proprietary systems (news.sap.com [8]). By moving to an Iceberg-based open table format, SAP is effectively saying that enterprises can no longer afford to have their data in walled gardens. The goal is a single, federated data layer where every analytics and AI tool has consistent, governed access to the full spread of enterprise information (news.sap.com [9]). This kind of unified architecture is poised to become a competitive necessity, as companies that can quickly integrate new data sources and maintain a holistic view of their business information will innovate faster and scale AI with far less friction.

References:
[1] www.ey.com — https://www.ey.com/en_au/insights/technology/the-next-digital-divide-are-you-among-the-ai-leaders-or-laggards#:~:text=emerging,your%20roadmap%20The%20rapidly%20changing
[2] news.sap.com — https://news.sap.com/2026/05/sap-to-acquire-dremio-unify-sap-and-non-sap-data-power-agentic-ai/#:~:text=Unify%20SAP%20and%20Non,expand%20SAP%20Business%20Data%20Cloud%E2%80%99s
[3] news.sap.com — https://news.sap.com/2026/05/sap-to-acquire-dremio-unify-sap-and-non-sap-data-power-agentic-ai/#:~:text=Unify%20SAP%20and%20Non,expand%20SAP%20Business%20Data%20Cloud%E2%80%99s
[4] www.dremio.com — https://www.dremio.com/blog/what-forresters-2026-data-lakehouse-landscape-signals-about-the-market-and-where-dremio-fits/#:~:text=lakehouses%20are%20now%20the%20default,Lakehouses
[5] news.sap.com — https://news.sap.com/2026/05/sap-to-acquire-dremio-unify-sap-and-non-sap-data-power-agentic-ai/#:~:text=Unify%20SAP%20and%20Non,expand%20SAP%20Business%20Data%20Cloud%E2%80%99s
[6] news.sap.com — https://news.sap.com/2026/05/sap-to-acquire-dremio-unify-sap-and-non-sap-data-power-agentic-ai/#:~:text=With%20Dremio%2C%20SAP%20Business%20Data,This%20means
[7] news.sap.com — https://news.sap.com/2026/05/sap-to-acquire-dremio-unify-sap-and-non-sap-data-power-agentic-ai/#:~:text=With%20Dremio%2C%20SAP%20Business%20Data,This%20means
[8] news.sap.com — https://news.sap.com/2026/05/sap-to-acquire-dremio-unify-sap-and-non-sap-data-power-agentic-ai/#:~:text=Unify%20SAP%20and%20Non,expand%20SAP%20Business%20Data%20Cloud%E2%80%99s
[9] news.sap.com — https://news.sap.com/2026/05/sap-to-acquire-dremio-unify-sap-and-non-sap-data-power-agentic-ai/#:~:text=With%20Dremio%2C%20SAP%20Business%20Data,This%20means

Real-Time Data and AI-Ready infrastructure

Another key differentiator emerging among AI leaders is the ability to harness data in real time. Many AI-driven services - from personalised customer experiences to predictive maintenance and automated decision-making - require immediate, up-to-the-minute data rather than static batches. Forward-looking enterprises are thus investing in streaming data pipelines and systems that feed their machine learning models with live data flows, so their AI can react to events as they happen. A prime example came with IBM's recent announcements: at its flagship conference this month, IBM's CEO stressed that companies winning in AI are not merely deploying more algorithms but "redesigning how their business operates," supported by a real-time, AI-ready data foundation (newsroom.ibm.com [1]) (newsroom.ibm.com [2]). To put weight behind this vision, IBM made one of its largest acquisitions since 2019 - a $11 billion deal to purchase Confluent, a leading cloud data streaming platform. By bringing Confluent's Kafka-based streaming technology into its watsonx data platform (now dubbed "IBM Confluent"), IBM is enabling enterprise AI systems to access fresh, trusted data from across disparate sources instantly, rather than waiting for overnight updates (www.stocktitan.net [3]).

In parallel, organisations are adopting new types of databases designed for AI's unique needs. Chief among these are vector databases, which store data (such as text, images, or customer records) as high-dimensional vectors to enable semantic search and retrieval for AI models. The booming interest in generative AI has turned vector stores from a niche tool into a centerpiece of modern AI infrastructure. In fact, Gartner projects that by 2026, more than 70% of all generative AI applications will rely on vector databases to handle unstructured data and provide context for AI queries (techspective.net [4]). This reflects a shift in enterprise data strategy: companies recognise that to get meaningful answers from AI (for example, a chatbot that can draw on all of a company's documents or real-time data), they need new data layers that can supply contextual knowledge at machine speed. Cloud and database vendors are racing to integrate vector search capabilities, and early adopters report significant gains in search relevance and faster AI-driven insights. The message is clear - AI-ready infrastructure now means streaming and vector-enabled data systems that ensure no

information is out-of-date or out-of-reach when the algorithm needs it.

References:

[1] newsroom.ibm.com — https://newsroom.ibm.com/2026-05-05-think-2026-ibm-delivers-the-blueprint-for-the-ai-operating-model-as-the-ai-divide-widens#:~:text=capabilities%20unveiled%20today%20address%20this,IBM%20is%20enabling%20organizations%20to
[2] newsroom.ibm.com — https://newsroom.ibm.com/2026-05-05-think-2026-ibm-delivers-the-blueprint-for-the-ai-operating-model-as-the-ai-divide-widens#:~:text=development%20give%20enterprises%20a%20unified,governance%20and%20sovereignty%20controls%20let
[3] www.stocktitan.net — https://www.stocktitan.net/news/IBM/think-2026-ibm-delivers-the-blueprint-for-the-ai-operating-model-as-ds1izli55ros.html#:~:text=match%20at%20L482%20data%20foundation,technologies%2C%20and%20pairing%20new%20capabilities
[4] techspective.net — https://techspective.net/2024/12/31/vector-databases-essential-tools-for-generative-ai-in-business/#:~:text=databases%20Gartner%20forecasts%20that%20by,The%20urgency%20stems

Proprietary Data: The new competitive moat

As artificial intelligence technology becomes more widespread, the competitive playing field is being defined by who owns the best data. In contrast to past years when owning a superior algorithm or model was seen as the key advantage, today's executives are increasingly treating proprietary data as invaluable intellectual property - the "moat" around their business. We're seeing real deals and partnerships formed for the primary purpose of securing unique data assets. At a recent industry gathering, experts highlighted how top AI firms are partnering with organisations that have rich, domain-specific datasets, underscoring that durable differentiation lives in the data underneath the model (marksbaughan.com [1]). For example, the AI company Anthropic has begun co-developing AI solutions using the internal knowledge repositories of global law firm Freshfields, in a multi-year collaboration to build AI-driven legal workflows (marksbaughan.com [2]). By training on Freshfields' vast trove of legal documents and expertise, Anthropic can create more specialised legal AI models while Freshfields gains early access to cutting-edge AI - a symbiotic arrangement built on a shared data moat.

Investors and markets are also recognising the outsized value of unique data. One analysis notes that companies with proprietary, high-quality training datasets are commanding 3 - 5 times higher valuations than competitors without those data assets (fourweekmba.com [3]). The rationale: as foundational AI models become commoditised or open-sourced, it's the exclusive data - customer interactions, domain-specific knowledge, real-time user behaviour, or industry-specific troves - that can't be easily replicated. Organisations that own such data can continually refine AI offerings in ways others cannot, leading to better insights and services. Conversely, firms that rely heavily on third-party or public data may find themselves at a strategic disadvantage. They risk having no unique training material to distinguish their AI, and worse, may inadvertently leak valuable information to the very platforms that could arm their competitors (www.welcome.ai [4]). This dynamic is encouraging more companies to keep a tight grip on their data exhaust and to invest in data collection, annotation, and stewardship as core competencies. In the AI era, your data isn't just fuelling your models - it is becoming the secret sauce that sets your business apart.

References:

[1] marksbaughan.com — https://marksbaughan.com/inta-2026-takeaways-data-moats-deal-flow-and-the-widening-valuation-gap/#:~:text=My%20three%20main%20takeaways%20are,The%20Munich
[2] marksbaughan.com — https://marksbaughan.com/inta-2026-takeaways-data-moats-deal-flow-and-the-widening-valuation-gap/#:~:text=My%20three%20main%20takeaways%20are,The%20Munich
[3] fourweekmba.com — https://fourweekmba.com/data-moats-the-ultimate-competitive-advantage-in-the-digital-age/#:~:text=2026%2C%20companies%20with%20proprietary%20training,processing%20capabilities%20have%20become%20the
[4] www.welcome.ai — https://www.welcome.ai/content/executives-emphasize-ai-sovereignty-amidst-global-data-control-risks#:~:text=towards%20data%20control%20and%20IP,financial%20performance%20and%20strategic%20positioning

Governance and sovereignty: The compliance mandate

The rise of AI as a business driver is being matched by a rise in regulatory and ethical scrutiny of data. Around the world, governments are introducing and refining laws to ensure AI developments don't compromise privacy, safety, or transparency (www.gdprregister.eu [1]). Europe in particular has been active: negotiators recently agreed on adjustments to the upcoming EU AI Act to give companies slightly more time to comply with requirements for "high-risk" AI, while accelerating transparency obligations (www.gdprregister.eu [2]). The message from regulators is that robust data governance is not optional. Companies will need to document data sources, ensure quality and fairness of training data, and maintain strict controls on data usage - or face hefty fines and legal challenges. Indeed, since GDPR came into effect, data privacy penalties in Europe have surged, totaling over €7.1 billion (with €1.2B issued in 2025 alone) (www.kiteworks.com [3]). Data mismanagement is evolving from a hidden technical debt into a direct compliance and business risk.

Leading enterprises are responding by making data sovereignty and governance a top C-suite priority. In many sectors - from finance to healthcare - companies are moving towards "sovereign cloud" solutions that keep sensitive data and AI processing confined within specific regions or jurisdictions to meet local laws and customer expectations (www.cxtoday.com [4]). Just last week, Workday announced a dedicated EU Sovereign Cloud aimed at European clients, ensuring that all their HR and financial data, along with any AI computations on that data, remain strictly in-region (www.cxtoday.com [5]). Such measures go hand-in-hand with building trust: customers and regulators alike want assurance that AI systems will not mishandle personal or critical data.

These efforts also align with a growing understanding that data governance is key to scaling AI responsibly. According to a global study by NTT Data, 35% of Chief AI Officers now say that enabling "private and sovereign AI" - AI that is built with protected data and runs in controlled environments - is their biggest barrier to adoption (www.helpnetsecurity.com [6]). Yet the same research indicates that organisations which make data and AI sovereignty a mission-critical priority are achieving up to 5 times higher returns on innovation and efficiency, and are 90% more likely to reach transformative AI outcomes (www.enterprisedb.com [7]) (www.enterprisedb.com [8]). In other words, the companies that proactively address data governance, privacy, and sovereignty are not just avoiding fines - they are positioning themselves for significantly better performance from their AI investments. In the coming months, we can expect more enterprises to follow the lead of these data-forward organisations, tightening data controls, investing in compliance-friendly architecture, and treating "data as law" as well as "data as an asset" in their AI strategies.

References:

[1] www.gdprregister.eu — https://www.gdprregister.eu/articles/eu-ai-act-update-new-deadlines/#:~:text=Priorities%20www,regulatory%20complexity%20and%20ease

[2] www.gdprregister.eu — https://www.gdprregister.eu/articles/eu-ai-act-update-new-deadlines/#:~:text=Just%20Changed,For

[3] www.kiteworks.com — https://www.kiteworks.com/gdpr-compliance/gdpr-fines-data-privacy-enforcement-2026/#:~:text=2026%20www,January%202026

[4] www.cxtoday.com — https://www.cxtoday.com/security-privacy-compliance/data-sovereignty-becomes-a-strategic-imperative-under-europes-compliance-rules/#:~:text=Localizing%20Data%20and%20AI%20Operations,Cloud%20gives%20our%20customers%20the

[5] www.cxtoday.com — https://www.cxtoday.com/security-privacy-compliance/data-sovereignty-becomes-a-strategic-imperative-under-europes-compliance-rules/#:~:text=Localizing%20Data%20and%20AI%20Operations,Cloud%20gives%20our%20customers%20the

[6] www.helpnetsecurity.com — https://www.helpnetsecurity.com/2026/05/19/ntt-sovereign-ai-strategy-report/#:~:text=Report%20A%20Playbook%20for%20Private,sovereign%20AI%20Private%20AI%20focuses

[7] www.enterprisedb.com — https://www.enterprisedb.com/press-releases/enterprises-prioritizing-sovereignty-across-data-and-agentic-and-genai-achieve-5x#:~:text=the%20findings%2C%20organizations%20that%20make,GenAI%20as%20their%20peers%2C%20demonstrating

[8] www.enterprisedb.com — https://www.enterprisedb.com/press-releases/enterprises-prioritizing-sovereignty-across-data-and-agentic-and-genai-achieve-5x#:~:text=higher%20ROI%20in%20terms%20of,While%20AI%20and%20data

Key Statistics

- Only 1 in 5 enterprise AI initiatives meet their key performance targets - the other 80% have been re-architected, paused, or abandoned, often due to lack of trusted, high-quality data ([securityboulevard.com](https://securityboulevard.com/2026/05/data-trust-ai-success-blog-series-why-most-ai-projects-are-failing/#:~:text=from%20MIND%E2%80%99s%20research%2C%20,KPIs%3F%20The%20same%20root%20cause))
- Lakehouses are now the default architecture for modern analytics and AI workloads, simplifying data integration and cost management (Forrester, Q1 2026) ([www.dremio.com](https://www.dremio.com/blog/what-forresters-2026-data-lakehouse-landscape-signals-about-the-market-and-where-dremio-fits/#:~:text=lakehouses%20are%20now%20the%20default,Lakehouses))
- Gartner predicts that by 2026, over 70% of generative AI applications will rely on vector databases for handling and retrieving data ([techspective.net](https://techspective.net/2024/12/31/vector-databases-essential-tools-for-generative-ai-in-business/#:~:text=databases%20Gartner%20forecasts%20that%20by,The%20urgency%20stems))
- Three-quarters of organisations admit their data governance has not kept pace with rapid AI adoption ([www.informatica.com](https://www.informatica.com/blogs/cdo-insights-2026-ai-adoption-accelerates-but-trust-and-governance-lag-behind.html#:~:text=of%20leaders%20still%20cite%20data,privacy%2C%20security%2C%20governance%2C%20and%20%E2%80%94))
- Organisations prioritising data/AI sovereignty achieve up to 5× higher ROI on innovation and efficiency, and are 90% more likely to achieve transformative AI results ([www.enterprisedb.com](https://www.enterprisedb.com/press-releases/enterprises-prioritizing-sovereignty-across-data-and-agentic-and-genai-achieve-5x#:~:text=the%20findings%2C%20organizations%20that%20make,GenAI%20as%20their%20peers%2C%20demonstrating)) ([www.enterprisedb.com](https://www.enterprisedb.com/press-releases/enterprises-prioritizing-sovereignty-across-data-and-agentic-and-genai-achieve-5x#:~:text=higher%20ROI%20in%20terms%20of,While%20AI%20and%20data))

KEY TAKEAWAY

AI success now depends more on data excellence than on model prowess. Recent developments show that leaders are investing in high-quality, well-governed, and unified data (often with real-time infrastructure and compliance in mind) to unlock AI's full business value, leaving less data-ready competitors behind.

Sources

Security Boulevard – Data Trust + AI Success: Why most AI projects are failing (May 20, 2026)
https://securityboulevard.com/2026/05/data-trust-ai-success-blog-series-why-most-ai-projects-are-failing/

SAP News – SAP to Acquire Dremio to Unify SAP and Non-SAP Data to Power Agentic AI (Press Release, May 4, 2026)
https://news.sap.com/2026/05/sap-to-acquire-dremio-unify-sap-and-non-sap-data-power-agentic-ai/

Dremio Blog – What Forrester's 2026 Data Lakehouse Landscape Signals About the Market (May 2026)
https://www.dremio.com/blog/what-forresters-2026-data-lakehouse-landscape-signals-about-the-market-and-where-dremio-fits/

IBM Newsroom – Think 2026: IBM Delivers the Blueprint for the AI Operating Model as the AI Divide Widens (May 5, 2026)
https://newsroom.ibm.com/2026-05-05-think-2026-ibm-delivers-the-blueprint-for-the-ai-operating-model-as-the-ai-divide-widens

Help Net Security – AI infrastructure is cracking under sovereignty demands (NTT DATA 2026 AI Report)
https://www.helpnetsecurity.com/2026/05/19/ntt-sovereign-ai-strategy-report/

CX Today – Data Sovereignty Becomes Strategic Business Priority in Europe (May 2026)
https://www.cxtoday.com/security-privacy-compliance/data-sovereignty-becomes-a-strategic-imperative-under-europes-compliance-rules/

Marks Baughan Insights – INTA 2026 Takeaways: Data Moats, Deal Flow, and the Widening Valuation Gap

(May 13, 2026)
https://marksbaughan.com/inta-2026-takeaways-data-moats-deal-flow-and-the-widening-valuation-gap/
FourWeekMBA – Data Moats: The Ultimate Competitive Advantage in the Digital Age (2026)
https://fourweekmba.com/data-moats-the-ultimate-competitive-advantage-in-the-digital-age/