LUMO INSIGHTS

Monday, 11 May 2026

# Data Strategy emerges as AI’s make-or-break factor

## Executive Summary

A wave of new reports and industry moves in the past week underscores that the key to unlocking AI’s business value lies in data readiness and architecture. From global surveys revealing massive gaps in data quality and governance, to major tech companies rolling out solutions for real-time integration and sovereign data control, the message is clearer than ever: enterprises must shore up their data foundations or risk falling behind in the AI era.

## AI adoption outpaces Data Readiness

A flurry of new research highlights a growing gulf between the rush to implement AI and the readiness of data to support it. A global survey by Dun & Bradstreet found that while an astonishing 97% of organisations now report active AI initiatives, only 5% describe their data as fully ready for AI (www.dnb.com [1]). Similarly, Cloudera’s recent Data Readiness Index reveals an emerging ‘AI readiness illusion’ - 96% of companies say they are integrating AI into core processes and 85% claim to have a clear data strategy, yet nearly four out of five admit that limited data access and quality issues still hinder their AI and analytics projects (www.cloudera.com [2]). In short, enthusiasm for AI is high, but basic data foundations are lagging behind.

Industry leaders warn that this misalignment is hobbling AI returns. According to an SAP announcement this week, 'most enterprise AI projects fail to deliver value not because of the AI itself, but because the underlying data is fragmented, locked in proprietary formats and stripped of the business context that makes it meaningful' (news.sap.com [3]). The result is a familiar pattern: many AI proofs-of-concept show promise but cannot scale or be trusted in production. Gartner research likewise indicates at least half of generative AI projects never make it past the pilot stage, largely due to poor data quality, inadequate risk controls, and unclear business value - factors unrelated to the sophistication of the algorithms (www.ibm.com [4]).

This data readiness crisis is driving a shift in executive perspective. The Chief Strategy Officer of Dun & Bradstreet noted this week that the constraint on AI is 'no longer the model; it is whether AI can operate on verified, continuously refreshed business data across systems' (www.dnb.com [5]). Without a solid “deterministic” data foundation - meaning clean, up-to-date, well-understood information - even the most advanced AI will produce only patchy insights and uneven ROI. The takeaway: to capture real business value from AI at scale, companies must first close their data readiness gap.

References:

[1] www.dnb.com — https://www.dnb.com/en-us/newsroom/press-releases/dnb-survey-finds-ai-impact-inflection-point.html#:~:text=Readiness%20Emerges%20as%20the%20Critical,released%20today%20by%20Dun

[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=striking%20paradox%3A%20while%2096,highlights%20an%20emerging%20%E2%80%9CAI%20readiness
[3] news.sap.com — https://news.sap.com/2026/05/sap-to-acquire-dremio-unify-sap-and-non-sap-data-power-agentic-ai/#:~:text=is%20still%20pending%20regulatory%20approval,familiar%20and%20costly%20pattern%3A%20pilots
[4] www.ibm.com — https://www.ibm.com/think/insights/why-most-enterprise-ai-projects-stall-before-scale#:~:text=least%2050,AI%20initiatives%20show%20strong%20early
[5] www.dnb.com — https://www.dnb.com/en-us/newsroom/press-releases/dnb-survey-finds-ai-impact-inflection-point.html#:~:text=holding%20organizations%20back,making.%22%20The

## Data quality & governance: breaking the bottleneck

As organisations dig into why AI pilots aren't turning into production wins, data quality and governance issues emerge as top culprits. Many enterprises still have data spread across incompatible databases, applications, and regions, making it hard for AI to get a consistent view of the business. One software provider remarked that 'most enterprise AI fails not because the models are flawed, but because the data is fragmented across disconnected systems and ungoverned' where AI agents operate (newsroom.servicenow.com [1]). When critical information is trapped in silos or lacks oversight, the result is often 'shallow intelligence' that can recommend actions but not reliably execute them in complex real-world processes (newsroom.servicenow.com [2]).

Recent surveys back up these concerns. Gartner finds that only 23% of IT leaders are very confident in their organisation's ability to manage data security and governance for AI, particularly for new generative AI tools (www.gartner.com [3]). And despite the proliferation of data initiatives, fewer than one in five companies (just 18%) say their data is fully governed and consistent across the organisation (www.cloudera.com [4]). This governance gap means AI systems often encounter ambiguous, low-quality data and unclear data ownership, leading to errors, bias, and compliance headaches that can derail projects.

To break these bottlenecks, forward-thinking firms are fortifying their data management practices. They are investing in data quality platforms, unified data catalogues, and master data management to reconcile definitions across the enterprise. Many are also appointing "data product" owners in different business domains - treating datasets as products with clear accountability for their quality, documentation, and lifecycle. Experts emphasise that creating AI-ready data is an ongoing discipline, not a one-off project (www.gartner.com [5]). By continuously cleaning, enriching, and governing data (and monitoring it for drifts or errors), companies ensure their AI models are always fed with accurate, up-to-date information.

References:
[1] newsroom.servicenow.com — 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#:~:text=siloed%20across%20systems%20that%20has,and%20applying%20a
[2] newsroom.servicenow.com — 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#:~:text=enterprise%20AI%20fails%20not%20because,workflow%20data%2C%20analytics%20insights%2C%20and
[3] www.gartner.com — https://www.gartner.com/en/newsroom/press-releases/2026-04-16-gartner-says-organizations-with-successful-ai-initiatives-invest-up-to-four-times-more-in-data-and-analytics-foundations#:~:text=Shift%205%3A%20Establish%20trust%20as,there%20is%20no%20value%20from
[4] 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
[5] 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=generation%20%28RAG%29%20integration%2C%20among%20others,ready%20data%20practice%20and

## Modern Data architecture: real-time & unified

Closing the data gap often requires rearchitecting how data is stored and delivered. Traditional enterprise data systems built for periodic reporting are evolving into more flexible "lakehouse" designs that can handle unstructured streams of information and deliver it in real time. One headline example this week: SAP announced plans to acquire Dremio, a leading data lakehouse platform, to help customers unify SAP and non-SAP datasets and run analytics and AI workloads in real time

(news.sap.com [1]). The move will make SAP’s Business Data Cloud an Apache Iceberg - powered lakehouse uniting previously siloed data, demonstrating a broader push across industry to eliminate fragmentation and to give AI models a single source of truth.

Enterprises are also embracing technologies that feed fresher data to AI systems. IBM, for instance, used its annual Think conference to highlight new streaming data capabilities (via a partnership with Confluent) to integrate real-time data into AI workflows (nand-research.com [2]). The rationale is clear: Many AI applications - especially “agentic” AI assistants that act autonomously - need up-to-the-second information to make decisions. Most legacy data pipelines were built for batch updates, so companies are investing in event-driven platforms and data fabrics to ensure their AI isn’t flying blind between weekly database updates.

Another piece of the next-generation data stack is the vector database, which has quickly risen in importance. These specialised databases store “vector embeddings” - complex numerical representations of text, images, and other content - enabling AI models to recall facts and context far more effectively. In 2026, vector databases have graduated from experimental tools to mission-critical infrastructure for AI leaders (www.marktechpost.com [3]). They serve as the memory layer underpinning AI-powered search and generative Q&A systems, allowing large language models to tap into a company’s proprietary knowledge base. Organisations that invest in the right data architecture - from real-time streaming to lakehouse platforms and vector stores - are finding that their AI solutions can operate with greater speed, scale, and relevance, giving them a leg up on competitors.

References:

[1] 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,more%20effectively%20run%20analytical%20and

[2] nand-research.com — https://nand-research.com/think-2026-ibms-ai-operating-model-takes-shape/#:~:text=Most%20enterprise%20data%20architectures%20were,utility%20of%20enterprise%20AI%20agents

[3] www.marktechpost.com — https://www.marktechpost.com/2026/05/10/best-vector-databases-in-2026-pricing-scale-limits-and-architecture-tradeoffs-across-nine-leading-systems/#:~:text=Pricing%2C%20Scale%20Limits%2C%20and%20Architecture,As

## Proprietary Data as the new moat

As cutting-edge AI models become widely accessible, companies are increasingly looking to their data as the source of competitive advantage. This was evident in a recent investor briefing where Thomson Reuters highlighted its troves of unique legal and tax data - combined with expert training and 'fiduciary-grade' AI models - as a key moat setting it apart from general-purpose AI providers (www.marketbeat.com [1]). That data-centric strategy coincides with business success: Thomson Reuters saw a 10% surge in first-quarter 2026 revenue, crediting the rapid adoption of its new AI-powered products (like the CoCounsel legal assistant and AI-enhanced research tools) for the boost (www.alphapilot.tech [2]). In effect, the company turned decades of proprietary content (bolstered by its 2023 acquisition of CaseText) into an AI-driven service that competitors without such data cannot easily match.

Similar dynamics are playing out across sectors. Financial institutions are leveraging exclusive troves of customer and market data to train AI models for insights that rivals can’t replicate. In the automotive realm, Tesla’s vehicles have logged over 8 billion self-driving miles, creating what one analyst dubs a 'data moat nobody can cross' in real-world driving information (tahaabbasi.com [3]). Even consumer platforms are walling off their data: last year Reddit began charging for API access and sued AI firms for scraping its content without consent, explicitly aiming to “capitalise on the value of its data” (www.socialmediatoday.com [4]). Whether it’s social media posts, proprietary research, or sensor data from millions of devices, companies now see data as proprietary intellectual property. The competitive logic is compelling - while AI algorithms can be bought or replicated, unique data cannot - and it is driving businesses to rethink data collection, partnerships, and monetisation

strategies.

References:

[1] www.marketbeat.com — https://www.marketbeat.com/instant-alerts/thomson-reuters-defends-ai-moat-as-compliance-complexity-fuels-growth-2026-05-09/#:~:text=and%20Westlaw%20Advantage%2C%20and%20arguing,and%20regulatory%20mandates%3B%20Thomson%20Reuters
[2] 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
[3] tahaabbasi.com — https://tahaabbasi.com/blog/taha-abbasi-tesla-fsd-8-billion-miles-milestone-autonomous-driving-2026#:~:text=Tesla%20FSD%20Surpasses%208%20Billion,highlights%20that%20Tesla%E2%80%99s%20Full%20Self
[4] www.socialmediatoday.com — https://www.socialmediatoday.com/news/reddit-launches-legal-action-against-ai-data-scrapers/803572/#:~:text=With%20Reddit%20looking%20to%20maximize,District%20Court%2C%20which%20seeks%20to

## Regulation & ethics: Data governance as a mandate

The push for better data practices isn't just coming from within companies - it's increasingly enforced from outside. Governments and regulators worldwide are writing new rules to rein in AI, with data at the centre of these policies. Europe's forthcoming AI Act, for example, will impose strict requirements on how AI systems are trained and deployed, including controls on data quality, transparency, and even a right for people to opt out of having their data used in AI training (www.gtlaw.com [1]). Under existing laws like GDPR, misuse of personal data by AI (for instance, scraping biometric or private content without proper consent) can already trigger fines as high as €20 /million or 4% of a company's global revenue (www.legalnodes.com [2]).

Recent enforcement actions show these aren't idle threats. In one high-profile case, Ireland's Data Protection Commission opened a sweeping inquiry into X (formerly Twitter) over its AI chatbot "Grok" generating non-consensual sexualized images, after allegations that the tool processed people's personal data to create deepfakes (www.politico.eu [3]). Likewise, earlier this year French authorities raided X's offices as part of an investigation into alleged misuse of personal data and other AI-related illegal content on the platform (time.com [4]). From Europe to the US and Asia, regulatory scrutiny is making data governance, privacy, and ethical AI usage top-of-mind issues for corporate boards.

Forward-looking tech providers are responding with solutions to help enterprises navigate this new landscape. At its Think 2026 conference last week, IBM unveiled 'Sovereign Core,' a software platform to help organisations build AI in "sovereign" cloud environments with full control over data location and compliance across hybrid infrastructures (newsroom.ibm.com [5]). In Japan, telecom giant SoftBank just launched a sovereign cloud in partnership with Oracle to power AI services while keeping all data and AI workloads within local data centres (blogs.oracle.com [6]). As one Gartner analyst aptly noted, 'without trust in the data, outputs and decisions of AI models and agents, there is no value from AI' (www.gartner.com [7]). In 2026 and beyond, ensuring that trust - through rigorous data governance, clear ownership, and respect for user rights - is now an essential part of any AI-ready data strategy.

References:

[1] www.gtlaw.com — https://www.gtlaw.com/en/insights/2024/7/eu-ai-acts-opt-out-trend-may-limit-data-use-for-training-ai-models#:~:text=the%20introduction%20of%20the%20EU,use%20in%20training%20AI%20models
[2] www.legalnodes.com — https://www.legalnodes.com/article/eu-ai-act-2026-updates-compliance-requirements-and-business-risks#:~:text=Risks%20www,of%20annual%20worldwide%20turnover
[3] www.politico.eu — https://www.politico.eu/article/ireland-inquiry-elon-musk-grok-privacy-artificial-intelligence-ai/#:~:text=Ireland%20launches%20%E2%80%98large,regulator%20said%20in%20a%20statement
[4] time.com — https://time.com/7366216/x-grok-offices-raided-france-united-kingdom-probe/#:~:text=billionaire%20issued%20a%20statement%20hours,%28Yaccarino%20stepped%20down%20from
[5] newsroom.ibm.com — https://newsroom.ibm.com/2026-05-05-think-2026-ibm-makes-digital-sovereignty-operational-with-general-availability-of-ibm-sovereign-core#:~:text=of%20IBM%20Sovereign%20Core%20Think,adoption%20accelerates%2C%20digital%20sovereignty%20has
[6] blogs.oracle.com — https://blogs.oracle.com/ai-and-datascience/whats-new-in-ai-may-2026#:~:text=3%20Nano%20Omni%20SoftBank%E2%80%99s%20Sovereign,strict%20data%20sovereignty%2C%20showing%20how
[7] www.gartner.com — https://www.gartner.com/en/newsroom/press-releases/2026-04-16-gartner-says-organizations-with-successful-ai-initiatives-invest-up-to-four-times-more-in-data-and-analytics-foundations#:~:text=Gartner%20survey%20of%20360%20IT,first%20D%26A%20leaders%20need%20to

## Key Statistics

- 97% of organisations have AI initiatives, but only 5% say their data is fully ready for AI ([www.dnb.com](https://www.dnb.com/en-us/newsroom/press-releases/dnb-survey-finds-ai-impact-inflection-point.html#:~:text=Readiness%20Emerges%20as%20the%20Critical,released%20today%20by%20Dun)).
- Companies with successful AI invest up to 4× more in data quality, governance, and training than those with poor AI outcomes ([www.gartner.com](https://www.gartner.com/en/newsroom/press-releases/2026-04-16-gartner-says-organizations-with-successful-ai-initiatives-invest-up-to-four-times-more-in-data-and-analytics-foundations#:~:text=Up%20to%20Four%20Times%20More,business%20and%20technology%20insights%20company)).
- Only 15% of enterprises are fully prepared for 'agentic AI' in production despite nearly 60% investing millions in it ([www.fivetran.com](https://www.fivetran.com/press/fivetran-launches-2026-agentic-ai-readiness-index-revealing-gap-between-enterprise-investment-and-data-preparedness-for-agentic-ai#:~:text=Only%2015,how%20prepared%20enterprise%20data%20environments)).

KEY TAKEAWAY

AI’s value hinges on data. Recent developments show that enterprises must urgently fix data quality issues, integrate silos, and enforce strong governance. Those who invest in robust, compliant data foundations now will outpace competitors in the AI race.

## Sources


Dun & Bradstreet Global Survey Finds AI Adoption Surging but Data Readiness Lags (Press Release, May 4, 2026)
https://www.dnb.com/en-us/newsroom/press-releases/dnb-survey-finds-ai-impact-inflection-point.html

Cloudera Press Release: Nearly 80% of Enterprises Say AI Is Held Back by Data Access Challenges (Apr 14, 2026)
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

SAP News: SAP to Acquire Dremio to Unify SAP and Non-SAP Data for 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/

ServiceNow News: Launch of Real-Time Data Foundation for Autonomous AI (Press Release, May 6, 2026)
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 News: Think 2026 – IBM Announces Sovereign Core for Digital Sovereignty (Press Release, May 5, 2026)
https://newsroom.ibm.com/2026-05-05-think-2026-ibm-makes-digital-sovereignty-operational-with-general-availability-of-ibm-sovereign-core

Fivetran Press: 2026 Agentic AI Readiness Index Shows Data Preparedness Gap (Press Release, May 5, 2026)
https://www.fivetran.com/press/fivetran-launches-2026-agentic-ai-readiness-index-revealing-gap-between-enterprise-investment-and-data-preparedness-for-agentic-ai

MarketBeat – Thomson Reuters Defends Data Moat Amid AI Growth (May 9, 2026)
https://www.marketbeat.com/instant-alerts/thomson-reuters-defends-ai-moat-as-compliance-complexity-fuels-growth-2026-05-09/

AlphaPilot: Thomson Reuters Q1 2026 Revenue Surges 10% on AI Success (Updated May 5, 2026)
https://www.alphapilot.tech/discover/thomson-reuters-q1-2026-revenue-surges-10-as-ai-integration-drives-organic-growth

AI Ireland: Proprietary Data as the New AI Moat (March 25, 2026)
https://aiireland.ie/2026/03/25/the-new-moat-why-proprietary-data-is-your-only-durable-competitive-advantage-in-ai/

Social Media Today: Reddit Sues to Protect User Data from AI Scraping (Oct 22, 2025)
https://www.socialmediatoday.com/news/reddit-launches-legal-action-against-ai-data-scrapers/803572/

Taha Abbasi Blog: Tesla FSD Surpasses 8 Billion Miles – The Data Moat Nobody Can Cross (Feb 19, 2026)
https://tahaabbasi.com/blog/taha-abbasi-tesla-fsd-8-billion-miles-milestone-autonomous-driving-2026

Politico EU: Irish DPC Launches Inquiry into X’s Grok Over AI-Generated Imagery (Feb 2026)
https://www.politico.eu/article/ireland-inquiry-elon-musk-grok-privacy-artificial-intelligence-ai/

TIME: Musk Blasts French Raid as 'Political': Deepfake Probe of X’s Grok AI (Feb 4, 2026)
https://time.com/7366216/x-grok-offices-raided-france-united-kingdom-probe/

Gartner Press Release: Successful AI Initiatives Invest 4x More in Data – Gartner Survey (Apr 16, 2026)
https://www.gartner.com/en/newsroom/press-releases/2026-04-16-gartner-says-organizations-with-successful-ai-initiatives-invest-up-to-four-times-more-in-data-and-analytics-foundations