LUMO INSIGHTS

Thursday, 11 June 2026

# Data strategy, not models, is driving today’s AI winners

## Executive Summary

A slew of new reports and industry moves this week all point to one conclusion: the key to unlocking AI’s value lies in data. From infrastructure and quality to governance, unique data assets and new tools, organisations that get their data strategy right are pulling ahead in the AI race.

## Data Infrastructure: The AI leader’s edge

Recent findings underscore how critical a strong data foundation is for AI success. In fact, a new survey revealed that only **6% of enterprise AI leaders feel their data infrastructure is fully ready for AI** (www.prnewswire.com [1]). This readiness gap - now seen as one of the biggest constraints on AI progress - highlights a simple truth: the companies reaping real benefits from AI are not necessarily those with the most advanced algorithms, but those with the best data foundations. As one data CEO put it, 'today, AI is constrained by data' (www.cdata.com [2]) rather than by model ingenuity.

The divide between AI leaders and laggards increasingly comes down to how they manage data. High-performing organisations almost always invest early in modern data architectures - from cloud data lakehouses to unified data warehouses - to ensure data is accessible, integrated and trustworthy. The **contrast is striking: 60% of companies at the highest level of AI maturity have built advanced data infrastructure, while 53% of those struggling with AI cite immature, fragmented data systems as a major hindrance** (www.cdata.com [3]). This gap is costing businesses time, money and competitive advantage (www.cdata.com [4]), as AI initiatives stall without the fuel of reliable data.

Accordingly, we’re seeing a strategic shift in IT investment. Organisations are recognising that improving data pipelines and platforms yields greater returns than obsessing over the latest machine-learning model. **Only 9% of companies now rank new AI model development as their top priority, while 83% are investing in centralised, consistent data access layers** (www.cdata.com [5]). By building "connected, contextual, semantically consistent" data environments (www.cdata.com [6]), these leaders create an AI-ready springboard - one that turns data into a durable asset rather than a bottleneck.

References:


[1] www.prnewswire.com — https://www.prnewswire.com/news-releases/cdata-study-finds-only-6-of-ai-leaders-believe-their-data-infrastructure-is-ready-for-ai-302629216.html#:~:text=infrastructure%20is%20fully%20ready%20for,and%20control%20are%20stalling%20AI

[2] www.cdata.com — https://www.cdata.com/company/press/state-of-ai-data-connectivity-report/#:~:text=constrained%20by%20models%20is%20over,Key%20Findings%3A%20AI%20teams%20are

[3] www.cdata.com — https://www.cdata.com/company/press/state-of-ai-data-connectivity-report/#:~:text=Infrastructure%20Gaps%20Hold%20Back%20AI,The%20gap%20is%20costing

[4] www.cdata.com — https://www.cdata.com/company/press/state-of-ai-data-connectivity-report/#:~:text=Infrastructure%20Gaps%20Hold%20Back%20AI,The%20gap%20is%20costing

[5] www.cdata.com — https://www.cdata.com/company/press/state-of-ai-data-connectivity-report/#:~:text=Only%209,It%27s%20determined%20by%20the
[6] www.cdata.com — https://www.cdata.com/company/press/state-of-ai-data-connectivity-report/#:~:text=constrained%20by%20models%20is%20over,Key%20Findings%3A%20AI%20teams%20are

## When Data falls short, so does AI

The flip side of this story is that poor data quality and silos are sabotaging AI projects at an alarming rate. Analysts have long warned that most AI initiatives don't fail due to algorithms, but due to subpar data. This week, that warning came with a stark statistic: **Gartner predicts that through 2026, 60% of AI projects lacking "AI-ready" data will be abandoned** (www.gartner.com [1]). Many companies enthusiastically pour resources into AI only to hit a wall when confronted with messy, unstructured, or inaccessible data. In banking, for example, an AI model is only as good as the customer and risk data it can learn from; if that data is erroneous or locked in departmental silos, the model's outputs won't be reliable.

New industry surveys illustrate these data pain points in detail. One report found that 74% of enterprises expect to integrate over 500 data sources to fuel their AI ambitions, creating significant complexity (www.fivetran.com [2]). Even heavily "centralised" organisations struggle: **67% of such companies still spend over 80% of their data engineering time just maintaining pipelines** - time siphoned away from innovation (www.fivetran.com [3]). And it's not just about the volume of data sources; timeliness matters too. **41% of organisations say the lack of real-time data access is preventing AI models from delivering timely insights**, while nearly a third point to data trapped in silos as a major blocker to AI success (www.fivetran.com [4]). In short, many teams are mired in assembling and cleaning data when they should be focusing on deploying AI solutions.

All this underscores that data quality is not a back-office IT concern - it's a fundamental business issue in the AI era. Every extra hour that data scientists spend on "data plumbing" is an hour not spent on value-adding analysis or innovation, a frustration echoed by 71% of AI teams who waste over a quarter of their time on data wrangling tasks (www.cdata.com [5]). Poor data also introduces risk: AI systems trained on flawed or biased data can produce erroneous or even unethical outcomes, exposing organisations to financial and reputational damage. The encouraging news is that improving data quality pays off. In fact, **61% of data leaders report that better data quality has already helped their generative AI pilots succeed** (www.informatica.com [6]), validating the push to invest in data preparation, master data management, and real-time integration. Leaders are learning that high-quality, well-governed data isn't a "nice to have" - it's the make-or-break factor for AI-driven ROI.

References:
[1] 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
[2] www.fivetran.com — https://www.fivetran.com/press/fivetran-report-finds-nearly-half-of-enterprise-ai-projects-fail-due-to-poor-data-readiness#:~:text=several%20key%20pain%20points%3A%2074,of
[3] www.fivetran.com — https://www.fivetran.com/press/fivetran-report-finds-nearly-half-of-enterprise-ai-projects-fail-due-to-poor-data-readiness#:~:text=several%20key%20pain%20points%3A%2074,of
[4] www.fivetran.com — https://www.fivetran.com/press/fivetran-report-finds-nearly-half-of-enterprise-ai-projects-fail-due-to-poor-data-readiness#:~:text=several%20key%20pain%20points%3A%2074,addressed%2C%20organizations%20will%20continue%20to
[5] www.cdata.com — https://www.cdata.com/company/press/state-of-ai-data-connectivity-report/#:~:text=drowning%20in%20data%20plumbing%3A%2071,%E2%80%94%20but%20still%20missing%3A%20100
[6] www.informatica.com — https://www.informatica.com/blogs/cdo-insights-2026-ai-adoption-accelerates-but-trust-and-governance-lag-behind.html#:~:text=impact%20everything%20from%20customer%20experience,governance%20isn%E2%80%99t%20optional%20%E2%80%94%20it%E2%80%99s

## The AI Governance & compliance gap

As AI expands across the enterprise, a new challenge has emerged: **how to govern and control AI deployments at scale**. A report from IBM's Institute for Business Value this week highlights a worrying accountability gap. After surveying 2,000 technology executives, IBM found that **two-thirds of CIOs and CTOs are being held accountable for AI systems they do not fully control** (newsroom.ibm.com [1]). Moreover, 70% of these tech leaders say that employees are now deploying tech (including AI and automation) faster than the IT department can keep track of it (newsroom.ibm.com [2]). It's no surprise, then, that **77% of organisations admit AI adoption is outpacing their governance and oversight capabilities** (newsroom.ibm.com [3]). In essence, AI is spreading into products and workflows faster than companies can implement policies to manage it responsibly.

Analysts are calling this mismatch "shadow AI" - a proliferation of AI systems operating outside the purview of central IT governance (www.cio.com [4]). AI-powered features are embedded in cloud applications, low-code tools, and departmental projects, which means innovation often happens at the fringes while oversight remains centralised. **CIOs retain responsibility for risk and compliance 'but lack control due to federated models and hyperscaler-led abstraction,'** as one Forrester analyst put it (www.cio.com [5]). Another expert noted that enterprises have *"decentralised experimentation far faster than they have decentralised accountability"*, leaving leadership with responsibility for AI outcomes without direct line of sight into every system leveraging sensitive data (www.cio.com [6]).

Closing this governance gap is now a top priority for C-level data leaders. Many are establishing enterprise-wide AI governance frameworks and cross-functional AI councils to regain visibility. Experts - and regulators - advise building **controls and compliance into AI systems from the start** rather than bolting them on later (www.prnewswire.com [7]). This approach pays dividends: organisations that bake automated governance and security checks into their AI platforms see *25% fewer incidents* when scaling up, compared to those relying on manual oversight (www.prnewswire.com [8]). And with **59% of enterprises citing regulatory compliance as the number-one challenge in managing data for AI** (www.fivetran.com [9]), robust governance isn't just about preventing accidents - it's essential for meeting data privacy laws and maintaining customer trust.

References:


[1] newsroom.ibm.com — https://newsroom.ibm.com/2026-06-08-new-ibm-study-finds-cios-and-ctos-face-growing-ai-control-gap-as-enterprise-deployment-scales#:~:text=IBM,tech%20CxOs%20anticipate%20a%2038

[2] newsroom.ibm.com — https://newsroom.ibm.com/2026-06-08-new-ibm-study-finds-cios-and-ctos-face-growing-ai-control-gap-as-enterprise-deployment-scales#:~:text=C,tech%20CxOs%20anticipate%20a%2038

[3] newsroom.ibm.com — https://newsroom.ibm.com/2026-06-08-new-ibm-study-finds-cios-and-ctos-face-growing-ai-control-gap-as-enterprise-deployment-scales#:~:text=CEO,%E2%80%9CFor%20CIOs%20and%20CTOs%2C%20the

[4] www.cio.com — https://www.cio.com/article/4182288/cios-are-being-held-accountable-for-ai-they-dont-fully-control-ibm-study-finds.html#:~:text=analyst%20at%20Greyhound%20Research,result%20is%20that%20technology%20leaders

[5] www.cio.com — https://www.cio.com/article/4182288/cios-are-being-held-accountable-for-ai-they-dont-fully-control-ibm-study-finds.html#:~:text=analysts%2C%20AI%20is%20increasingly%20being,result%20is%20that%20technology%20leaders

[6] www.cio.com — https://www.cio.com/article/4182288/cios-are-being-held-accountable-for-ai-they-dont-fully-control-ibm-study-finds.html#:~:text=unified%20AI%20governance%20model%2C%20so,developer%20tools%2C%20bypassing%20central%20IT

[7] www.prnewswire.com — https://www.prnewswire.com/news-releases/new-ibm-study-finds-cios-and-ctos-face-growing-ai-control-gap-as-enterprise-deployment-scales-302793417.html#:~:text=governance%20capabilities.%20,on%20manual%20governance%2C%20incident%20risk

[8] www.prnewswire.com — https://www.prnewswire.com/news-releases/new-ibm-study-finds-cios-and-ctos-face-growing-ai-control-gap-as-enterprise-deployment-scales-302793417.html#:~:text=confidence.,of%20tech%20CxOs

[9] www.fivetran.com — https://www.fivetran.com/press/fivetran-report-finds-nearly-half-of-enterprise-ai-projects-fail-due-to-poor-data-readiness#:~:text=enterprises%20allocate%20over%2080%20percent,ambition%20without%20execution%20is%20costing

## Data as a competitive moat

With AI models and tools increasingly accessible, businesses are discovering that **their proprietary data is becoming the key competitive differentiator**. A popular LinkedIn essay shared this week argues that traditional software advantages are eroding - features can be copied in days, UIs cloned in a weekend - due to accelerated AI-driven development (enterprisezone.cc [1]). For example, tasks that used to require five years of coding and millions of dollars can now be accomplished almost overnight with modern AI-assisted software generation (enterprisezone.cc [2]). This rapid commoditisation of technology means **the defensibility of pure software IP is declining**, pushing companies to find alternative ways to stand out.

The consensus? The *only true, lasting AI advantage may lie in unique data*. As the LinkedIn author pointed out, *'the one thing AI can’t fake is years of real, verified, proprietary data'* (enterprisezone.cc [3]). No matter how sophisticated an AI model is, it can’t simply conjure up a rich trove of authentic customer interactions, transaction records, or domain-specific knowledge that a business has accumulated over time. You 'can’t prompt your way to 300 million professional profiles' or generate fully accurate, up-to-date product data out of thin air (enterprisezone.cc [4]). In other words, competitors might access similar models, but they cannot easily recreate your data assets.

Companies that have invested in building these data “moats” are already reaping the benefits. A striking example comes from financial services: Bloomberg’s new GPT model was trained on **40 years of proprietary financial data, news, and research** that no rival can easily replicate (opag.io [5]). Analysts note that *“the data moat is not the model architecture - any competitor could build a similar model - but the dataset that powers it”* (opag.io [6]). This exclusive data foundation gives Bloomberg a persistent edge. Likewise, across industries, leading firms are treating high-quality data as a form of intellectual property - safeguarding it, enriching it, and leveraging it to fuel AI solutions that competitors can’t match.

References:


[1] enterprisezone.cc — https://enterprisezone.cc/the-data-moat-why-proprietary-information-is-the-new-competitive-edge-according-to-federicodona/#:~:text=rapid%20acceleration%20of%20product%20development,The%20core
[2] enterprisezone.cc — https://enterprisezone.cc/the-data-moat-why-proprietary-information-is-the-new-competitive-edge-according-to-federicodona/#:~:text=rapid%20acceleration%20of%20product%20development,The%20core
[3] enterprisezone.cc — https://enterprisezone.cc/the-data-moat-why-proprietary-information-is-the-new-competitive-edge-according-to-federicodona/#:~:text=capabilities,Data%20as%20the
[4] enterprisezone.cc — https://enterprisezone.cc/the-data-moat-why-proprietary-information-is-the-new-competitive-edge-according-to-federicodona/#:~:text=datasets,to%20Federicodonatone%2C%20this%20unique%20characteristic
[5] opag.io — https://opag.io/insights/ai-competitive-moats-valuing-proprietary-systems#:~:text=data%20through%20their%20daily%20use,research%20that%20advances%20beyond%20publicly
[6] opag.io — https://opag.io/insights/ai-competitive-moats-valuing-proprietary-systems#:~:text=data%20through%20their%20daily%20use,research%20that%20advances%20beyond%20publicly


## Building AI-Ready Data architectures

The tech industry is responding to these challenges with new platforms and architectural approaches that put data at the centre. Last week, at its Build 2026 conference, Microsoft unveiled **Project Rayfin**, an open-source toolkit designed to help developers deploy AI-powered applications on its unified Microsoft Fabric data platform (www.infoworld.com [1]). Rayfin aims to make it vastly easier to launch AI apps by combining data storage, databases, APIs, and machine learning runtime in one managed environment (www.infoworld.com [2]). By eliminating much of the "data plumbing" and integration work typically needed to stand up an AI solution, Rayfin not only speeds development but also inherently keeps **application data within a governed, analytics-ready estate** (www.infoworld.com [3]) (www.infoworld.com [4]). This "governance by default" approach - with security, compliance, and access controls automatically enforced from day one - serves as what one

expert called a 'governed on-ramp' for enterprise AI, helping prevent the sprawl of unapproved, siloed AI projects (www.infoworld.com [5]) (www.infoworld.com [6]).

Data platform leaders are also doubling down on delivering **built-in business context and governance**. At its annual user conference, Snowflake introduced *Horizon Context*, a semantic metadata layer that curates and connects business definitions, data lineage, and access rules across the company's Data Cloud (www.cio.com [7]). By providing a **centralised, governed map of the entire data estate**, this tool gives AI systems and analytics teams a consistent understanding of key metrics and terms (for instance, what counts as a "customer" or "revenue") when they access data (www.cio.com [8]). In the past, many enterprises created patchworks of data catalogues, BI definitions, and governance tools - resulting in different teams using conflicting data definitions (www.cio.com [9]). With a unified context layer, Snowflake aims to ensure an AI agent answering a customer query or generating a report uses the same trusted data and rules that a human analyst would - reducing errors and confusion.

Beyond these specific products, the broader trend is towards **AI-ready data infrastructure** that seamlessly blends flexibility, speed, and control. One fast-emerging technology is the vector database, which stores data in a format that generative AI models can query to retrieve facts, documents, and context in real time. Gartner forecasts that by 2026, **over 30% of enterprises will have adopted vector databases to ground their AI models with proprietary business data - up from less than 2% in 2023** (aibuzz.blog [10]). This explosion in *retrieval-augmented generation (RAG)* techniques shows how quickly companies are moving to connect their AI systems with the **latest, most relevant internal data**. Likewise, modern "data lakehouse" architectures - which combine the scale of data lakes with the reliability and governance of data warehouses - are becoming the default for large enterprises. In an open lakehouse, a **unified metadata catalogue acts as the central brain**, enforcing consistent security policies and data definitions across all analytics and AI workloads (www.architectureandgovernance.com [11]). The takeaway: to enable AI at scale, organisations are redesigning their data ecosystems for agility (through cloud and streaming data pipelines), for trust (through governance and data quality controls), and for strategic value (through proprietary data assets).

References:


[1] www.infoworld.com — https://www.infoworld.com/article/4181166/rayfin-signals-microsofts-push-to-make-fabric-an-ai-app-runtime.html#:~:text=Rayfin%20signals%20Microsoft%E2%80%99s%20push%20to,assisted%20development%2C%20writing%20code%20is
[2] www.infoworld.com — https://www.infoworld.com/article/4181166/rayfin-signals-microsofts-push-to-make-fabric-an-ai-app-runtime.html#:~:text=consultant%20David%20Linthicum%2C%20Rayfin%20increases,that%20addresses%20the%20risks%20CIOs
[3] www.infoworld.com — https://www.infoworld.com/article/4181166/rayfin-signals-microsofts-push-to-make-fabric-an-ai-app-runtime.html#:~:text=consultant%20David%20Linthicum%2C%20Rayfin%20increases,that%20addresses%20the%20risks%20CIOs
[4] www.infoworld.com — https://www.infoworld.com/article/4181166/rayfin-signals-microsofts-push-to-make-fabric-an-ai-app-runtime.html#:~:text=data%20closer%20to%20the%20analytics,by%20a%20coordinated%20network%20of
[5] www.infoworld.com — https://www.infoworld.com/article/4181166/rayfin-signals-microsofts-push-to-make-fabric-an-ai-app-runtime.html#:~:text=data%20closer%20to%20the%20analytics,by%20a%20coordinated%20network%20of
[6] www.infoworld.com — https://www.infoworld.com/article/4181166/rayfin-signals-microsofts-push-to-make-fabric-an-ai-app-runtime.html#:~:text=added,ramp%2C%E2%80%9D
[7] www.cio.com — https://www.cio.com/article/4180170/snowflakes-horizon-context-aims-to-give-ai-agents-a-common-understanding-of-the-business.html#:~:text=systems%20are%20only%20as%20reliable,across%20AI%20and%20analytics%20systems
[8] www.cio.com — https://www.cio.com/article/4180170/snowflakes-horizon-context-aims-to-give-ai-agents-a-common-understanding-of-the-business.html#:~:text=Context%20should%20reduce%20operational%20complexity,models%2C%20security%20policies%2C%20and%20custom
[9] www.cio.com — https://www.cio.com/article/4180170/snowflakes-horizon-context-aims-to-give-ai-agents-a-common-understanding-of-the-business.html#:~:text=and%20business%20context%20needed%20to,would%20end%20up%20with%20slightly
[10] aibuzz.blog — https://aibuzz.blog/embeddings-vector-databases-explained/#:~:text=%E2%9C%85Gartner%20forecasts%20that%20over%2030,primary%20driver%20of%20vector%20database
[11] www.architectureandgovernance.com — https://www.architectureandgovernance.com/artificial-intelligence/breaking-the-silos-the-rise-of-the-open-lakehouse-architecture-in-2026/#:~:text=Governance%20and%20Metadata%20Layer%3A%20A,data%20definitions%20remain%20consistent%20and

## Key Statistics

- Only 6% of enterprise AI leaders say their data infrastructure is fully AI-ready ([www.prnewswire.com](https://www.prnewswire.com/news-releases/cdata-study-finds-only-6-of-ai-leaders-believe-their-data-infrastructure-is-ready-for-ai-302629216.html#:~:text=infrastructure%20is%20fully%20ready%20for,and%20control%20are%20stalling%20AI)).
- Gartner predicts 60% of AI projects will be abandoned by 2026 due to lack of “AI-ready” data ([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)).
- 67% of CIOs & CTOs are being held accountable for AI systems they do not fully control ([newsroom.ibm.com](https://newsroom.ibm.com/2026-06-08-new-ibm-study-finds-cios-and-ctos-face-growing-ai-control-gap-as-enterprise-deployment-scales#:~:text=IBM,tech%20CxOs%20anticipate%20a%2038)).
- Over 30% of enterprises will use vector databases by 2026 (vs <2% in 2023) to integrate proprietary data into AI models ([aibuzz.blog](https://aibuzz.blog/embeddings-vector-databases-explained/#:~:text=%E2%9C%85Gartner%20forecasts%20that%20over%2030,primary%20driver%20of%20vector%20database)).
- 83% of companies are investing in centralised data access layers for AI, while only 9% still prioritise new model development ([www.cdata.com](https://www.cdata.com/company/press/state-of-ai-data-connectivity-report/#:~:text=Only%209,It%27s%20determined%20by%20the)).

KEY TAKEAWAY

The case for AI is now a case for data. New research and real-world examples show that robust data foundations, governance, and unique proprietary information are what separate successful AI initiatives from costly failures and compliance nightmares.

### Sources


CData Study Finds Only 6% of AI Leaders Believe Their Data Infrastructure Is Ready for AI
https://www.prnewswire.com/news-releases/cdata-study-finds-only-6-of-ai-leaders-believe-their-data-infrastructure-is-ready-for-ai-302629216.html

New IBM Study Finds CIOs and CTOs Face Growing AI Control Gap as Enterprise Deployment Scales
https://newsroom.ibm.com/2026-06-08-new-ibm-study-finds-cios-and-ctos-face-growing-ai-control-gap-as-enterprise-deployment-scales

Rayfin signals Microsoft’s push to make Fabric an AI app runtime
https://www.infoworld.com/article/4181166/rayfin-signals-microsofts-push-to-make-fabric-an-ai-app-runtime.html

The Data Moat: Why Proprietary Information is the New Competitive Edge, According to Federicodonatone…
https://enterprisezone.cc/the-data-moat-why-proprietary-information-is-the-new-competitive-edge-according-to-federicodona/

Lack of AI-Ready Data Puts AI Projects at Risk
https://www.gartner.com/en/newsroom/press-releases/2025-02-26-lack-of-ai-ready-data-puts-ai-projects-at-risk

Fivetran Report Finds Nearly Half of Enterprise AI Projects Fail Due to Poor Data Readiness
https://www.fivetran.com/press/fivetran-report-finds-nearly-half-of-enterprise-ai-projects-fail-due-to-poor-data-readiness

CDO Insights 2026 – AI Adoption Accelerates, But Trust and Governance Lag Behind
https://www.informatica.com/blogs/cdo-insights-2026-ai-adoption-accelerates-but-trust-and-governance-lag-behind.html

Snowflake’s Horizon Context aims to give AI agents a common understanding of the business
https://www.cio.com/article/4180170/snowflakes-horizon-context-aims-to-give-ai-agents-a-common-understanding-of-the-business.html