LUMO INSIGHTS

Monday, 4 May 2026

Data strategy drives AI Winners - And leaves others behind

Executive Summary

Enterprise leaders are learning that the success of AI initiatives hinges on a factor often overlooked: data strategy and readiness. This week's developments - from cross-cloud data tools to stark industry failure stats and regulatory debates - all highlight how data architecture, quality, and governance determine whether AI projects thrive or fail.

The data foundation divide

The gulf between AI leaders and laggards is increasingly traced to data foundations. New industry analysis shows that only 28% of enterprise AI initiatives currently deliver a meaningful return on investment (ROI). Meanwhile, 20% of projects fail outright, and a staggering 57% of tech managers have experienced at least one AI project collapse - wasting an estimated $486 billion annually on unrealised AI promises.

These failures often have little to do with the AI models themselves. In fact, 77% of AI project failures are caused by organisational shortcomings rather than technical glitches. The underlying culprits include disjointed data architectures and unclear ownership of data and processes, which leave advanced analytics built on shaky ground. As one senior executive put it, the road to AI heaven goes through data hell (techcrunch.com [1]) - a reminder that without fixing data issues, even the best algorithms falter.

Forward-thinking companies are responding by doubling down on data infrastructure and governance from the start. They're shifting from "vibe-based" AI spending to measurable business value, from fragmented experiments to a unified data strategy, and from launching projects without metrics to a governance-first approach. Organisations that build dedicated data infrastructure teams and embed governance early see up to 3 - 5× higher returns - along with 30 - 50% lower costs - on their AI initiatives.

References:

[1] techcrunch.com — https://techcrunch.com/2025/05/07/servicenow-acquires-data-world-two-months-after-acquiring-moveworks/#:~:text=their%20data%20%E2%80%9CAI,enterprises%20organize%20and%20easily%20search

Quality, governance & ownership - the new bottlenecks

Poor data quality and siloed information remain the biggest blockers to AI success. Only about one-third of businesses report meaningful progress in AI adoption, and chief data officers (CDOs) overwhelmingly blame subpar data quality (www.itpro.com [1]). Over two-thirds (68%) of CDOs now say that poor data quality is their top challenge in harnessing AI, far outpacing issues like talent gaps (www.itpro.com [2]). Essentially, without high-quality data, AI just multiplies mistakes faster (www.weforum.org [3]).

Data governance and clear ownership are equally crucial. If no one is accountable for data, even a well-built AI can end up unloved and unused. Recent findings show 41% of failed AI projects floundered because they were deployed with no business owner to take responsibility for their adoption. Another 34% failed by solving the wrong problem - a sign that AI efforts must be tightly aligned with business needs from day one.

The good news is that organisations treating data as a strategic asset are overcoming these hurdles. In leading enterprises, governance isn't viewed as red tape but as a catalyst for innovation. Experts observe that winning companies foster a culture of treating data like a product, where quality and trust are paramount, and "governance doesn't slow progress; it makes it sustainable" (www.cio.com [4]). With strong stewardship and unified standards, data turns from an obstacle into fuel for AI.

A powerful success story comes from the finance sector. Morgan Stanley curated a digital library of 100,000 internal documents and fine-tuned a GPT-4 wealth management assistant so it could "effectively answer any question" from that massive corpus (openai.com [5]). Thanks to this robust data foundation and rigorous oversight, the firm reports over 98% adoption of the AI assistant by its financial advisors (openai.com [6]) - a testament that trusted data and governance drive real business value from AI.

References:

[1] www.itpro.com — https://www.itpro.com/business/data-and-insights/ai-projects-are-faltering-as-cdos-grapple-with-poor-data-quality#:~:text=please%20log%20in,in

[2] www.itpro.com — https://www.itpro.com/business/data-and-insights/ai-projects-are-faltering-as-cdos-grapple-with-poor-data-quality#:~:text=quality.%20More%20than%20two,critical%20factor%20when%20leveraging%20data

[3] www.weforum.org — https://www.weforum.org/stories/2026/01/why-data-readiness-is-now-a-strategic-imperative-for-businesses/#:~:text=outputs%20often%20get%20blamed%20on,How%20to

[4] www.cio.com — https://www.cio.com/article/4083265/why-80-of-ai-projects-fail-and-how-smart-enterprises-are-finally-getting-it-right.html#:~:text=match%20at%20L83%20governance%20doesn%E2%80%99t,related.%20But%20the%20most

[5] openai.com — https://openai.com/index/morgan-stanley/#:~:text=expanding%20document%20library,%E2%80%9D

[6] openai.com — https://openai.com/index/morgan-stanley/#:~:text=has%20been%20really%20impactful.%E2%80%9D%2098,adoption%20in%20wealth

Proprietary data as the new AI moat

Companies are realising their proprietary data can be a durable competitive moat in the AI era. At the HumanX 2026 summit in San Francisco, AI leaders argued that proprietary data - not choice of model - now determines competitive advantage (www.msn.com [1]). In other words, choosing between algorithms like GPT-4 or Claude is less important than having unique, high-quality data to feed them.

Real moves in the market back up this view. Mastercard, which processes billions of transactions globally, has built a generative AI model fuelled by its enormous trove of payment data (www.pymnts.com [2]). Plaid, whose platform connects to thousands of banks, similarly unveiled an 'intelligent finance' model trained on its vast financial records (www.pymnts.com [3]). These payment networks boast one-of-a-kind datasets that give them a "data moat" - a protected edge no general AI provider can easily replicate (www.pymnts.com [4]). By leveraging years of proprietary information, they aim to unlock insights and services that competitors without similar data simply can't match.

A similar pattern is emerging across industries. Healthcare organisations are aggregating patient and clinical data to train AI-driven diagnostics, while manufacturers are tapping into IoT sensor data to optimise production with AI. Many firms are also acquiring data management and analytics companies to gain access to rich data sources (consider ServiceNow's purchase of data.world to bolster its AI readiness last year) (techcrunch.com [5]). The message is clear: in the age of generative AI, a company's data foundation itself is becoming its most critical intellectual property.

References:

[1] www.msn.com — https://www.msn.com/en-us/news/other/humanx-2026-panel-declares-proprietary-data-the-real-ai-moat/gm-GML3F1BA01#:~:text=declares%20proprietary%20data%20the%20real,quality%2C%20relevant%20data
[2] www.pymnts.com — https://www.pymnts.com/artificial-intelligence-2/2026/fintechs-race-to-build-foundation-models-on-proprietary-data/#:~:text=Mastercard%20announced%20in%20March%20that,scale%20AI%20system
[3] www.pymnts.com — https://www.pymnts.com/artificial-intelligence-2/2026/fintechs-race-to-build-foundation-models-on-proprietary-data/#:~:text=Mastercard%20announced%20in%20March%20that,scale%20AI%20system
[4] www.pymnts.com — https://www.pymnts.com/artificial-intelligence-2/2026/fintechs-race-to-build-foundation-models-on-proprietary-data/#:~:text=normalization%2C%20categorization%20and%20risk%20signaling,advantage%20no%20LLM%20creator%20has
[5] techcrunch.com — https://techcrunch.com/2025/05/07/servicenow-acquires-data-world-two-months-after-acquiring-moveworks/#:~:text=PDT%20%C2%B7%20May%207%2C%202025,Ventures%2C%20Prologis%20Ventures%2C%20and%20Shasta

Modern data architectures for an AI-Ready enterprise

The data architecture landscape is changing rapidly to meet AI's growing needs. One major trend is the push to eliminate silos with federated, cross-cloud data platforms. Rather than forcing teams to copy data into one system, cloud vendors now enable analytics and AI to run where the data already lives. Google's new cross-cloud lakehouse, for example, allows organisations to analyse data across different cloud providers without needing to move it (blog.google [1]). Similarly, IBM just announced a "zero-copy" data federation feature in its watsonx.data platform that lets companies query external data sources without duplication (community.ibm.com [2]). By leveraging open table formats and distributed processing, these approaches maintain a single source of truth and reduce the cost and risk of moving sensitive data around.

Enterprises are also seeking real-time data pipelines to keep AI models up-to-the-second. As more customer interactions and decisions are driven by AI, streaming data infrastructure is becoming essential for AI readiness. New "real-time lakehouse" designs combine traditional data lakes with streaming SQL engines to deliver sub-second queries on continually updated information (risingwave.com [3]). This means AI-powered services - from personalised marketing to autonomous operations - can react instantly to the latest events, giving businesses a competitive time advantage.

Even core databases are evolving for the AI age. The once-specialised vector databases (designed to store AI embeddings for semantic search) are turning into mainstream features. In fact, by 2026 every major cloud and database provider - from AWS and Microsoft to MongoDB - has added native vector search support (dev.to [4]). This shift lets companies enrich existing data platforms with AI-driven search and recommendation capabilities without adding new, complex systems, streamlining deployment and reducing friction.

References:

[1] blog.google — https://blog.google/innovation-and-ai/infrastructure-and-cloud/google-cloud/google-cloud-next-26-recap/#:~:text=also%20includes%20Cross,As%20cybersecurity
[2] community.ibm.com — https://community.ibm.com/community/user/blogs/bijo-alex-thomas/2026/04/20/watsonxdata-news-a-technical-touchpoint-april-2026#:~:text=Zero,maintaining%20a%20single%20source%20of
[3] risingwave.com — https://risingwave.com/blog/data-lakehouse-architecture-2026/#:~:text=processes%20events%20continuously%20using%20SQL%2C,database%20as%20part%20of%20a
[4] dev.to — https://dev.to/actiandev/whats-changing-in-vector-databases-in-2026-3pbo#:~:text=a%20growing%20movement%20from%20cloud,Deployment%20limitations%20add%20another%20critical

Navigating data regulation and ethics

Shifting regulations and ethical considerations around data are now integral to AI strategy. In Europe, policymakers are debating ways to ease data privacy rules to fuel AI innovation. Draft proposals would dial back parts of GDPR to simplify data sharing for AI development (www.dw.com [1]) - a move intended to cut red tape and help European firms compete globally. While businesses may cheer a lighter regulatory touch, privacy advocates warn that loosening data protections could undermine public trust.

In the U.S., meanwhile, companies face a growing patchwork of data laws. Without a single federal privacy standard, 19 states now have their own comprehensive data privacy regulations (citipen.com [2]). This fragmentation makes it challenging to scale AI solutions nationwide without a meticulous compliance strategy. On top of that, dedicated AI regulations are looming; for instance, the EU's AI Act will start enforcing new transparency and data governance mandates by August 2026. The takeaway for CTOs and CDOs is that robust data governance and ethical management are no longer optional. Leaders must treat compliance as a core pillar of AI readiness - both to avoid legal pitfalls and to build the trust that makes sustainable AI success possible.

References:

[1] www.dw.com — https://www.dw.com/en/eu-plans-to-ease-gdpr-laws-and-ai-constraints-in-major-shift/a-74792773#:~:text=against%20harvesting%20user%20data%20for,number%20of%20data%20protection%20rules
[2] citipen.com — https://citipen.com/ai-regulation-data-privacy-laws-2026-explained/#:~:text=Enforcement%20Frontier%20With%20no%20comprehensive,2024%20with%20a%20focus%20on

Key Statistics

- 68% of Chief Data Officers say poor data quality is their top AI challenge ([www.itpro.com](https://www.itpro.com/business/data-and-insights/ai-projects-are-faltering-as-cdos-grapple-with-poor-data-quality#:~:text=quality.%20More%20than%20two,critical%20factor%20when%20leveraging%20data)).
- 60% of AI projects without AI-ready data will be abandoned by 2026 ([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=their%20AI%20efforts,Gartner%2C%20to%20discuss%20how%20AI)).
- Over 98% of Morgan Stanley's advisor teams use its internal GPT-4 assistant daily ([openai.com](https://openai.com/index/morgan-stanley/#:~:text=has%20been%20really%20impactful.%E2%80%9D%2098,adoption%20in%20wealth)).

KEY TAKEAWAY

Leading companies are doubling down on data integration, quality and governance - and they're pulling ahead in AI. Those that don't fix data foundations and compliance now risk big spending on AI with little to show for it.

Sources

ServiceNow acquires Data.World months after snatching up Moveworks
https://techcrunch.com/2025/05/07/servicenow-acquires-data-world-two-months-after-acquiring-moveworks/
AI projects are faltering as CDOs grapple with poor data quality
https://www.itpro.com/business/data-and-insights/ai-projects-are-faltering-as-cdos-grapple-with-poor-data-quality
Why 80% of AI projects fail — and how smart enterprises are finally getting it right
https://www.cio.com/article/4083265/why-80-of-ai-projects-fail-and-how-smart-enterprises-are-finally-getting-it-right.html
Why data readiness is now a strategic imperative for businesses
https://www.weforum.org/stories/2026/01/why-data-readiness-is-now-a-strategic-imperative-for-businesses/
Proprietary Data Gives Payments Firms a New AI Moat
https://www.pymnts.com/artificial-intelligence-2/2026/fintechs-race-to-build-foundation-models-on-proprietary-data/
HumanX 2026 panel declares proprietary data the real AI moat
https://www.msn.com/en-us/news/other/humanx-2026-panel-declares-proprietary-data-the-real-ai-moat/gm-GML3F1BA01
7 highlights and announcements from Google Cloud Next '26
https://blog.google/innovation-and-ai/infrastructure-and-cloud/google-cloud/google-cloud-next-26-recap/
watsonx.data News | A Technical Touchpoint April 2026
https://community.ibm.com/community/user/blogs/bijo-alex-thomas/2026/04/20/watsonxdata-news-a-technical-touchpoint-april-2026
Data Lakehouse Architecture in 2026: Streaming, Iceberg, and the Real-Time Layer
https://risingwave.com/blog/data-lakehouse-architecture-2026/

What’s Changing in Vector Databases in 2026
https://dev.to/actiandev/whats-changing-in-vector-databases-in-2026-3pbo
AI Laws 2026: What New Regulations Mean for You
https://citipen.com/ai-regulation-data-privacy-laws-2026-explained/
Morgan Stanley uses AI evals to shape the future of financial services
https://openai.com/index/morgan-stanley/