LUMO INSIGHTS

Monday, 10 August 2026

Data strategy: The make-or-break factor for AI success

Executive Summary

This week's developments emphasise that advanced AI is only as effective as the data behind it. New studies and regulatory moves show that top-performing companies are pulling ahead by investing heavily in data quality, governance, and modern architecture, while those that neglect their data foundations face failed AI projects and compliance troubles. From Europe's tough new data rules to real-time "lakehouse" platforms, the message is clear: businesses must treat their data as a strategic asset to unlock AI's value.

AI leaders vs laggards: The Data divide

At a high level, a new divide is emerging in the business world between companies successfully harnessing AI and those falling behind - and the difference often comes down to data. New research from Hitachi Vantara underscores this point: weak data foundations are preventing 58% of surveyed organisations in the US and Canada from realising value from their AI investments (www.hitachivantara.com [1]), contributing to an estimated $108 /billion in wasted global AI spending per year. In contrast, organisations that have built "data-mature" environments with modern infrastructure, integrated data, and strong governance are far more likely to see returns. In fact, 84% of these data leaders report measurable ROI on AI projects, compared to only 48% of their data-laggard peers (www.hitachivantara.com [2]).

What sets these AI leaders apart is a strategic focus on data. Research by Gartner finds that enterprises successfully scaling AI invest up to four times more (as a percentage of revenue) in data quality, data governance, and related talent than their struggling peers (www.webpronews.com [3]). This extra investment in data pays off - mature data management setups yield about 65% greater improvements in revenue and cost savings from AI initiatives (www.webpronews.com [4]). The highest performers ensure their data is clean, interconnected, and readily accessible to data science teams; they don't just pour money into advanced algorithms without first fortifying the data foundation beneath.

These leading organisations also treat data itself as a core strategic asset. In an IBM global survey, 78% of Chief Data Officers said that leveraging proprietary data for competitive advantage is now a top priority in their AI strategy (newsroom.ibm.com [5]). In practice, AI leaders are building robust internal data platforms, breaking down silos and curating high-quality datasets unique to their business. This proprietary data - data that competitors don't have - is increasingly seen as key to maintaining an edge as AI technologies spread across industries.

References:

[1] www.hitachivantara.com — https://www.hitachivantara.com/en-us/news/gl260127#:~:text=enterprises%20race%20to%20adopt%20AI%2C,IT%20leaders%20across%2015%20countries

[2] www.hitachivantara.com — https://www.hitachivantara.com/en-us/news/gl260127#:~:text=the%20value%20of%20AI%20initiatives,say%20AI%20is
[3] www.webpronews.com — https://www.webpronews.com/ais-silent-force-quadruple-investments-in-data-core-separate-winners-from-laggards/#:~:text=thriving%20AI%20projects%20spend%20up,Data%20quality%20ensures%20inputs%20are
[4] 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
[5] newsroom.ibm.com — https://newsroom.ibm.com/2025-11-13-ibm-study-chief-data-officers-redefine-strategies-as-ai-ambitions-outpace-readiness#:~:text=Readiness%20IBM%20Study%3A%20Chief%20Data,The%20findings

AI Project failures: The Data factor

For all the excitement around AI, a sobering truth has emerged: most AI projects fail to deliver on their promise, and data issues are frequently to blame. Analysts estimate that over 80% of enterprise AI initiatives ultimately fall short of their expected goals (www.greencodesoftware.com [1]). For example, RAND Corporation found that more than 80% of AI projects provided no measurable business value, and an MIT study reported that 95% of organisations experimenting with generative AI saw zero tangible return on those pilots (www.greencodesoftware.com [2]).

The cause of these failures is usually not the AI technology itself, but rather the quality and readiness of the data feeding it. As one in-depth industry review concluded, “the root cause, in the overwhelming majority of cases, isn’t the algorithm - it’s the data” (www.greencodesoftware.com [3]), with 85% of failed AI projects citing poor data quality as the primary culprit (www.greencodesoftware.com [4]). Common problems include training data that is incomplete, inaccurate or biased, and data pipelines that cannot deliver relevant, up-to-date information to AI systems. In essence, even the most sophisticated model will stumble if it’s trained on unreliable or fragmented data.

This data deficit is derailing AI initiatives at an alarming rate. Gartner forecasts that through 2026, 60% of AI projects lacking “AI-ready” data will be abandoned before they ever reach production (www.greencodesoftware.com [5]). In many cases, projects start strong but then hit a wall when moving from lab to launch: in 2025 alone, 42% of companies had to scrap at least one AI initiative due to data quality or integration issues, more than double the rate of the previous year (www.greencodesoftware.com [6]).

The story of one financial institution’s failed fraud detection initiative is a cautionary example. The company built a promising model that achieved 94% accuracy on a carefully curated historical dataset. But when the AI was deployed on live data in production, the model’s accuracy plummeted to 67% and it began flagging thousands of false positives per day - overwhelming the compliance team and undermining trust in the system. After just a few months, the project was cancelled (www.beri.net [7]). The culprit was a gap in data quality and consistency: the real-world data flowing through the firm’s legacy systems was far messier and more variable than the pristine data used to train the model.

The flip side is that when AI projects do succeed, high-quality data is often the hero. 59% of organisations report that using clean, well-governed data was the top driver of their successful AI initiatives (www.hitachivantara.com [8]). The takeaway for senior leaders is clear: improving data quality, access, and governance isn’t just an IT chore - it’s essential to turning AI investments into real business value.

References:

[1] www.greencodesoftware.com — https://www.greencodesoftware.com/en/blog/ai-projects-fail-data-readiness-gartner-2026#:~:text=practices%20needed%20to%20support%20AI,least%20one%20AI%20initiative%20in
[2] www.greencodesoftware.com — https://www.greencodesoftware.com/en/blog/ai-projects-fail-data-readiness-gartner-2026#:~:text=practices%20needed%20to%20support%20AI,least%20one%20AI%20initiative%20in
[3] www.greencodesoftware.com — https://www.greencodesoftware.com/en/blog/ai-projects-fail-data-readiness-gartner-2026#:~:text=technology%20works,management%20practices%20needed%20for%20AI
[4] www.greencodesoftware.com — https://www.greencodesoftware.com/en/blog/ai-projects-fail-data-readiness-gartner-2026#:~:text=and%2085,solution%20isn%27t%20buying%20more%20technology

[5] www.greencodesoftware.com — https://www.greencodesoftware.com/en/blog/ai-projects-fail-data-readiness-gartner-2026#:~:text=should%20be%20hanging%20in%20every,RAND%20Corporation%20found
[6] www.greencodesoftware.com — https://www.greencodesoftware.com/en/blog/ai-projects-fail-data-readiness-gartner-2026#:~:text=S%26P%20Global%20reported%20that%2042,The%20pattern%20is%20consistent%3A%20the
[7] www.beri.net — https://www.beri.net/article/ai-project-failure-complete-guide-2026#:~:text=Real%20example%3A%20A%20financial%20services,Project%20canceled%20after%204%20months
[8] www.hitachivantara.com — https://www.hitachivantara.com/en-us/news/gl260127#:~:text=performance%20and%20return%20on%20investment,quality%20data%20%E2%80%93%20rising%20to

AI regulations raise the Data bar

This week also brought a turning point in how governments are compelling companies to get their data in order for AI. On August 2, 2026, the European Union's landmark AI Act entered its enforcement phase, meaning firms deploying high-risk AI systems (e.g. in lending or hiring) must now meet strict data transparency and governance requirements. Organisations using these "high-risk" AI applications are required to document exactly what data was used to train and test their algorithms, to ensure compliance with quality standards and to enable oversight of potential biases.

Regulators aren't just asking for paperwork - they expect technical proof. Under the new EU rules, companies can be required to show auditors the specific datasets that trained an AI model, how those data were vetted for bias or errors, and even the precise model version and logic that contributed to a given automated decision. Meeting such demands will test many companies' data architectures. An organisation can no longer simply point to a data lake and claim compliance; it needs rigorous data lineage, version control and audit trails for every input and output an AI system produces.

The immediate impact in Europe has been a scramble for compliance. France's data protection authority (CNIL), for example, has already ordered 14 financial institutions to produce the required technical documentation for their AI-based credit scoring systems - and pointedly denied extensions for firms that were unprepared. Meanwhile, regulatory pressure is mounting globally: in the U.S., federal AI legislation is stalled, leaving states like California and Colorado to impose their own AI data rules; in China, regulators have already fined 12 companies a total of ¥4.2 million in the first week of enforcing new AI data regulations.

It's no surprise, then, that ensuring data governance and "explainability" in AI systems has become a top priority for C-suites. Enterprises are establishing cross-functional data and AI governance committees and investing in privacy-centric data architectures to meet the rising standards. The goal is not only to avoid penalties, but to build the trust needed for AI to truly deliver. As one expert warns, "Without trust in the data, outputs and decisions of AI models and agents, there is no value from AI" (www.webpronews.com [1]). In other words, robust data oversight has become essential to achieving any business value from AI.

References:
[1] www.webpronews.com — https://www.webpronews.com/ais-silent-force-quadruple-investments-in-data-core-separate-winners-from-laggards/#:~:text=security%20and%20governance%20for%20generative,The%20Governance%20Overhaul%20Imperative%20This

Modernising Data architecture for AI

Amid these pressures, enterprises are upgrading their information infrastructure to be "AI-ready." One approach gaining momentum is the data "lakehouse" - a new type of platform that combines the scale of data lakes with the governance and performance of data warehouses. In a blog post this week summarising Forrester's latest assessment of enterprise data platforms, analysts noted that the role of the lakehouse is expanding from a passive repository of information to an active "execution layer" for AI applications (www.forrester.com [1]). In short, these systems are becoming the real-time data backbones on which AI models and AI-driven products run.

Designing data architecture for AI requires building in trust, speed, and openness. As AI applications increasingly make autonomous decisions, any weakness in data quality or oversight can turn into an execution risk. Forrester emphasises that next-generation platforms must have strong controls “built-in” - such as automated data lineage, continuous data quality checks, and fine-grained access controls - so AI systems can reliably train on and act on the data with minimal human intervention (www.forrester.com [2]). In other words, the data platform is no longer just a backend store; it’s a critical part of the AI product pipeline, ensuring that data feeding AI is timely, accurate, and secure.

Flexibility is another priority. Companies need the freedom to plug in various models and tools over time, so modern data architectures are increasingly built with open standards to avoid vendor lock-in. Lakehouse platforms that support open file formats, interoperable metadata, and standard APIs allow organisations to adapt as new AI technologies emerge (www.forrester.com [3]). Another emerging capability is “vector” search, which lets AI applications retrieve information based on semantic similarity. Leading data platforms now integrate vector database technology to support these needs, enabling AI systems (like chat-based assistants) to fetch relevant text or images from enterprise knowledge in milliseconds (www.forrester.com [4]).

The rapid adoption of such technologies underscores how central data architecture has become to AI success. Gartner forecasts that by 2026, more than 30% of new enterprise applications using generative AI will rely on vector databases to find and retrieve information - a huge leap from less than 5% of such applications in 2023 (www.devx.com [5]). Decisions about data platforms have thus become strategic choices for the C-suite: picking the right mix of data warehouse, lakehouse, streaming and vector database solutions can determine how quickly a company can scale promising AI prototypes into business-ready solutions.

References:

[1] www.forrester.com — https://www.forrester.com/blogs/takeaways-from-the-forrester-wave-data-lakehouses-q3-2026/#:~:text=lakehouse%20is%20becoming%20an%20execution,data%20quality%2C%20governance%2C%20lineage%2C%20or
[2] www.forrester.com — https://www.forrester.com/blogs/takeaways-from-the-forrester-wave-data-lakehouses-q3-2026/#:~:text=autonomous%20decisions%2C%20weaknesses%20in%20data,access%20controls%2C%20continuous%20data%20quality
[3] www.forrester.com — https://www.forrester.com/blogs/takeaways-from-the-forrester-wave-data-lakehouses-q3-2026/#:~:text=monitoring%2C%20and%20policy%20enforcement%20as,AI
[4] www.forrester.com — https://www.forrester.com/blogs/takeaways-from-the-forrester-wave-data-lakehouses-q3-2026/#:~:text=enablement%20is%20the%20new%20competitive,Interested%20in
[5] www.devx.com — https://www.devx.com/uncategorized/vector-databases-beyond-ai-hype-2026/#:~:text=integrate%20it%20well,DevX%20previously

Data as a competitive moat

As AI technology becomes more democratised, companies are increasingly leaning on proprietary data - rather than proprietary algorithms - to hold a competitive edge. The cost of advanced AI capabilities is plummeting; for instance, as of early 2026 the average cost of running a powerful billion-parameter AI model has fallen by 94% since 2023 (authority-journal.com [1]). When any competitor can access similar algorithms, the differentiator shifts to the data those algorithms are trained on.

For many executives, this means reimagining business data as proprietary intellectual property. An IBM study of 1,700 global organisations found that 78% of Chief Data Officers now prioritise leveraging proprietary data to help differentiate their companies in the market (newsroom.ibm.com [2]). These unique troves of customer, operational, and industry-specific information are increasingly seen as high-ground that competitors can’t easily replicate. The more a company can inject its own data into AI models - from fine-tuning large language models to feeding real-time decision engines - the more it can build AI capabilities that others cannot readily copy.

There is growing evidence that these data advantages translate into lasting performance gains. A Stanford-led study of 312 enterprises found that companies with unified, high-quality data infrastructures experienced 67% less degradation in their AI systems' performance over 18 months compared to firms with siloed, inconsistent data (authority-journal.com [3]). Even as cutting-edge models became available to all, the gap between data-rich and data-poor organisations actually widened - better models amplified the benefits of better data rather than closing the difference (authority-journal.com [4]). In other words, simply having access to powerful AI is not enough; what matters most is having the best data to feed it.

It's a lesson leading AI adopters have already internalized. As AI pioneer Andrew Ng bluntly puts it, 'The model is infrastructure. The data is the business.' (authority-journal.com [5]) In practice, enterprises that double down on building unique, high-quality data assets - and continuously refining them - are creating durable competitive moats. In the AI era, a well-curated dataset that no rival possesses can be a more sustainable advantage than any single breakthrough in algorithms.

References:
[1] authority-journal.com — https://authority-journal.com/the-proprietary-data-moat-who-actually-wins-when-ai-gets-cheap/#:~:text=In%20the%20first%20quarter%20of,the%20moat%20cannot%20be%20built
[2] newsroom.ibm.com — https://newsroom.ibm.com/2025-11-13-ibm-study-chief-data-officers-redefine-strategies-as-ai-ambitions-outpace-readiness#:~:text=Readiness%20IBM%20Study%3A%20Chief%20Data,The%20findings
[3] authority-journal.com — https://authority-journal.com/the-proprietary-data-moat-who-actually-wins-when-ai-gets-cheap/#:~:text=The%20Stanford%20HAI%20Enterprise%20AI,widened%2C%20not%20narrowed%2C%20as%20model
[4] authority-journal.com — https://authority-journal.com/the-proprietary-data-moat-who-actually-wins-when-ai-gets-cheap/#:~:text=feedback%20loops%20from%20AI%20outputs,not%20compensate%20for%20worse%20data
[5] authority-journal.com — https://authority-journal.com/the-proprietary-data-moat-who-actually-wins-when-ai-gets-cheap/#:~:text=applied%20AI%20research%2C%20has%20been,%E2%80%9D

Key Statistics

- 58% of surveyed North American firms saw no value from AI due to weak data foundations, contributing to $108 billion in wasted global AI investment per year ([www.hitachivantara.com](https://www.hitachivantara.com/en-us/news/gl260127#:~:text=enterprises%20race%20to%20adopt%20AI%2C,IT%20leaders%20across%2015%20countries)).
- 84% of data-mature organisations report measurable AI ROI, compared with only 48% of data laggards ([www.hitachivantara.com](https://www.hitachivantara.com/en-us/news/gl260127#:~:text=performance%20and%20return%20on%20investment,quality%20data%20%E2%80%93%20rising%20to)).
- Gartner predicts that through 2026, organisations will abandon 60% of AI projects that lack 'AI-ready' data ([www.greencodesoftware.com](https://www.greencodesoftware.com/en/blog/ai-projects-fail-data-readiness-gartner-2026#:~:text=should%20be%20hanging%20in%20every,RAND%20Corporation%20found)).
- 59% of companies say using high-quality data was the top driver of their successful AI projects ([www.hitachivantara.com](https://www.hitachivantara.com/en-us/news/gl260127#:~:text=performance%20and%20return%20on%20investment,quality%20data%20%E2%80%93%20rising%20to)).
- By 2026, more than 30% of new enterprise applications using generative AI will rely on vector databases, up from under 5% in 2023 ([www.devx.com](https://www.devx.com/uncategorized/vector-databases-beyond-ai-hype-2026/#:~:text=integrate%20it%20well,DevX%20previously)).

KEY TAKEAWAY

Even the best AI can't succeed without quality, well-governed data. To capture AI's benefits and avoid failures or fines, data leaders must double down on data foundations - from modern infrastructure to unique, trusted data assets that rivals can't easily replicate.

Sources

Legacy Data Infrastructure Is Driving $108 Billion in Annual Wasted AI Investment, New Survey Finds
https://www.prnewswire.com/news-releases/legacy-data-infrastructure-is-driving-108-billion-in-annual-wasted-ai-investment-new-survey-finds-302670840.html

AI's Silent Force: Quadruple Investments in Data Core Separate Winners from Laggards
https://www.webpronews.com/ais-silent-force-quadruple-investments-in-data-core-separate-winners-from-laggards/

Gartner Predicts 60% of AI Projects Will Be Abandoned Due to Poor Data Readiness — And LATAM Has an Even Bigger Problem
https://www.greencodesoftware.com/en/blog/ai-projects-fail-data-readiness-gartner-2026

Takeaways From The Forrester Wave™: Data Lakehouses, Q3 2026
https://www.forrester.com/blogs/takeaways-from-the-forrester-wave-data-lakehouses-q3-2026/

The Proprietary Data Moat: Who Actually Wins When AI Gets Cheap
https://authority-journal.com/the-proprietary-data-moat-who-actually-wins-when-ai-gets-cheap/

80% of AI Projects Fail: The Hidden Cause
https://www.beri.net/article/ai-project-failure-complete-guide-2026

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

Vector Databases in 2026: Moving Beyond the AI Hype
https://www.devx.com/uncategorized/vector-databases-beyond-ai-hype-2026/