LUMO INSIGHTS

Monday, 1 June 2026

AI Success Hinges on Data Strategy, Not Just Models

Executive Summary

This week’s top developments reveal that the key to winning with AI lies in data strategy and infrastructure. From cloud giants retooling for AI to new warnings on data quality and emerging regulations, the message is clear: robust data foundations are the deciding factor for AI readiness.

Rearchitecting Data Infrastructure for AI

Enterprises are learning that yesterday's IT infrastructure can't handle tomorrow's AI-driven workloads. Cloud architecture long optimized for human-scale interactions is being redesigned for AI. For example, Amazon Web Services (AWS) just launched a next-generation OpenSearch service, effectively a hybrid search and vector database, designed to support agentic AI workloads (techcrunch.com [1]). This system can dynamically scale up to handle bursts of activity when AI agents spin up hundreds of queries, then scale back down when idle (techcrunch.com [2]), a clear response to how AI usage patterns differ from traditional web traffic.

The surge in machine-generated traffic is forcing a broader rethink of enterprise networks. Cloudflare reports that bots (including AI crawlers and assistants) already account for 31% of internet traffic (techcrunch.com [3]). It's no surprise that 97% of IT leaders now consider modern, AI-ready networks and data infrastructure critical for deploying AI, and 91% have increased their network investments accordingly (newsroom.cisco.com [4]). Organizations on the leading edge are upgrading data pipelines, storage, and compute capacity to ensure their platforms can deliver real-time insights and handle AI’s intense, spiky workloads. Those that don't risk performance bottlenecks and even outages as AI demands scale up.

References:

[1] techcrunch.com — https://techcrunch.com/2026/05/28/the-internet-is-being-rebuilt-for-machines/#:~:text=arrived,agents%20still%20represent%20a%20relatively

[2] techcrunch.com — https://techcrunch.com/2026/05/28/the-internet-is-being-rebuilt-for-machines/#:~:text=Serverless%2C%20a%20fully%20managed%20search,of

[3] techcrunch.com — https://techcrunch.com/2026/05/28/the-internet-is-being-rebuilt-for-machines/#:~:text=increasingly%20populated%20by%20agents,bot%20requests%20during%20that%20period

[4] newsroom.cisco.com — https://newsroom.cisco.com/c/r/newsroom/en/us/a/y2025/m06/cisco-research-a-major-infrastructure-shift-is-underway-ai-could-double-the-strain-or-solve-it.html#:~:text=Leaders%20believe%20the%20right%20infrastructure,%E2%80%93

Data Quality & Governance: AI's True Bottleneck

While AI tools grow more powerful, many companies find themselves held back not by algorithms but by data shortcomings. A recent industry survey revealed that 42% of enterprises have seen over half of their AI projects delayed, underperform, or fail due to data readiness issues (internalaudit360.com [1]). Likewise, at this week's Data Summit 2026 in Boston, enterprise data architect Milan Parikh argued that the real bottleneck in AI initiatives is almost always the data foundation, not the model (www.ibtimes.com [2]). In other words, if the data architecture under an AI project isn't robust,

integrated, and high-quality, even the most sophisticated algorithms will struggle to deliver value.

The costs of poor data practices are enormous. Parikh cited research showing 60-70% of data teams run duplicate data pipelines in different departments, and that when organizations bolt on governance after the fact, their time-to-insight can be 3-4x longer (www.ibtimes.com [3]). Moreover, poor data quality costs large enterprises an average of $12.9 million per year (www.ibtimes.com [4]). These inefficiencies and hidden costs mean that without clean, well-governed data, many AI initiatives never progress beyond the pilot stage.

The good news is that forward-thinking data leaders are tackling these issues head-on. Approaches like the medallion architecture (which organizes data into Bronze, Silver, and Gold quality tiers) are being adopted as blueprints for AI-ready data pipelines (www.ibtimes.com [5]). By enforcing data quality checks, common standards, and governance at each stage, this layered strategy ensures that by the time information reaches the Gold layer, it is fully trusted and analytics-ready. Parikh demonstrated how an end-to-end platform like Microsoft's new Fabric can streamline this process by unifying data lakes, real-time streaming, transformations, and governance tools in one place (www.ibtimes.com [6]). His core advice for executives: build governance and quality into the data architecture from day one, rather than treating it as an afterthought (www.ibtimes.com [7]).

References:

[1] internalaudit360.com — https://internalaudit360.com/poor-data-readiness-is-plaguing-corporate-ai-projects/#:~:text=stuck%20managing%20infrastructure%20instead%20of,of%20enterprises%20say%20regulatory%20compliance

[2] www.ibtimes.com — https://www.ibtimes.com/inside-data-foundation-problem-behind-enterprise-ai-failure-milan-parikh-takes-case-data-summit-3803265#:~:text=priority,Fellow%20of%20the%20British%20Computer

[3] www.ibtimes.com — https://www.ibtimes.com/inside-data-foundation-problem-behind-enterprise-ai-failure-milan-parikh-takes-case-data-summit-3803265#:~:text=are%20difficult%20to%20dismiss,of%20that%2C%20you%20have%20already

[4] www.ibtimes.com — https://www.ibtimes.com/inside-data-foundation-problem-behind-enterprise-ai-failure-milan-parikh-takes-case-data-summit-3803265#:~:text=to,a%20new%20copy%20of%20the

[5] www.ibtimes.com — https://www.ibtimes.com/inside-data-foundation-problem-behind-enterprise-ai-failure-milan-parikh-takes-case-data-summit-3803265#:~:text=with%20quality%20gates%20at%20every,for%20operational%20and%20IoT%20workloads

[6] www.ibtimes.com — https://www.ibtimes.com/inside-data-foundation-problem-behind-enterprise-ai-failure-milan-parikh-takes-case-data-summit-3803265#:~:text=Microsoft%20Fabric%20as%20the%20Delivery,the%20Gold%20analytics%20layer%20with

[7] www.ibtimes.com — https://www.ibtimes.com/inside-data-foundation-problem-behind-enterprise-ai-failure-milan-parikh-takes-case-data-summit-3803265#:~:text=DirectLake%20connectivity%2C%20eliminating%20the%20data,for%20enterprises%20starting%20from%20scratch

Avoiding a New Wave of AI Silos

Another emerging challenge is the risk of fragmented AI efforts across the enterprise. As generative AI and personal AI agents become more pervasive, individual teams are deploying their own solutions, often without a coordinated data strategy. Industry analysts warn that if every person and department builds its own isolated AI systems, companies will end up recreating the same old data silos and inconsistencies even faster, with greater operational risk (thecuberesearch.com [1]). This scenario is reminiscent of the early PC era, when uncontrolled tech adoption led to fragmented information across organizations – but with AI, the stakes and speed of sprawl are even higher.

To counter this, leading organizations are establishing unified data and AI platforms to serve as a central 'System of Intelligence' for the business. The idea is to organize enterprise knowledge, data, and business logic in one place where AI applications and agents can access trusted information and context (thecuberesearch.com [2]). With a unified architecture, companies can allow innovation at the edges (letting teams experiment with new AI tools) while ensuring all systems draw from a common, well-governed data source. This prevents the proliferation of conflicting metrics and results, keeping everyone on the same page.

The race to provide these integrated platforms is heating up. Cloud data leaders like Snowflake and Databricks, for example, are rapidly expanding from analytics into full-stack AI enablement

(thecuberesearch.com [3]). Both are enhancing their platforms – from data lakehouse storage to built-in machine learning and vector search capabilities – aiming to become the go-to backbone for enterprise AI. Tech giants and startups alike recognize that whichever platform becomes a company's de facto System of Intelligence will secure a deep competitive moat. For CIOs and CDOs, the mandate is clear: avoid a patchwork of ad hoc AI solutions by investing in an integrated data architecture that can support enterprise-wide AI growth.

References:

[1] thecuberesearch.com — https://thecuberesearch.com/316-breaking-analysis-personal-agents-light-the-fuse-as-snowflake-and-databricks-move-up-the-ai-stack/#:~:text=documents%2C%20spreadsheets%20and%20dashboards,the%20new%20front%20end%20%E2%80%93
[2] thecuberesearch.com — https://thecuberesearch.com/316-breaking-analysis-personal-agents-light-the-fuse-as-snowflake-and-databricks-move-up-the-ai-stack/#:~:text=of%20Intelligence%20%28SoI%29,longer%20just%20data%20platforms%20serving
[3] thecuberesearch.com — https://thecuberesearch.com/316-breaking-analysis-personal-agents-light-the-fuse-as-snowflake-and-databricks-move-up-the-ai-stack/#:~:text=readable%20and%20eventually%20executable%20by,version%20of%20this%20control%20point

Data as a Competitive Moat

All these developments underscore that data itself is turning into the most critical competitive asset in the AI era. As cutting-edge models become widely accessible, proprietary datasets – and the ability to harness them – are emerging as a primary source of sustained advantage (ai.via.news [1]). Many experts argue that enterprise AI competition is shifting from who has the most advanced algorithms to who has the best data. Established companies with years of domain-specific data have a head start; they can train AI systems on troves of historical, high-quality information that new entrants simply don't possess (ai.via.news [2]).

This realization is changing how businesses approach data ownership and sharing. Organizations are increasingly treating their data as vital intellectual property and a strategic moat. Many are now cautious about sending sensitive information to external AI providers, fearing they might inadvertently give away their 'secret sauce' to competitors. Instead, firms are exploring ways to develop and fine-tune AI models using their own data within secure environments. By closely guarding and optimizing their unique data resources, companies can build AI capabilities that rivals cannot easily replicate, securing a long-term edge.

References:

[1] ai.via.news — https://ai.via.news/enterprise-ai-infrastructure/operational-data-not-model-capability-is-the-new-enterprise-ai-moat#:~:text=Enterprise%20AI%20competition%20is%20shifting,competitive%20outcomes%20will%20be%20decided
[2] ai.via.news — https://ai.via.news/enterprise-ai-infrastructure/operational-data-not-model-capability-is-the-new-enterprise-ai-moat#:~:text=Enterprise%20AI%20competition%20is%20shifting,operation%20is%20where%20competitive%20outcomes

Regulation: Navigating Data, Privacy and AI

Finally, global regulators are shaping the data strategies of AI-driven businesses. This week, the European Commission advanced proposals to adjust data laws to better accommodate AI innovation (iapp.org [1]). For example, planned updates to the EU's GDPR would explicitly permit companies to process personal data for AI under a 'legitimate interests' justification - provided strong privacy safeguards are in place (iapp.org [2]). The goal of these Digital Omnibus reforms is to spur AI development while upholding individuals' rights. At the same time, regulators are sharpening their focus on data security and ethics. Misuse of personal information by AI systems – for instance in biometric or emotion recognition – can already trigger fines up to €20 million or 4% of global turnover under GDPR rules (www.legalnodes.com [3]).

The upshot for data and technology leaders is that compliance must go hand in hand with innovation. Leading organizations are working closely with legal teams to anticipate how emerging AI regulations (from the EU's AI Act to data localization requirements) will impact their data architecture and cross-

border data flows. By investing in strong governance, transparency, and privacy-preserving techniques now, companies can pursue AI opportunities confidently and competitively, without running afoul of evolving regulations.

References:
[1] iapp.org — https://iapp.org/news/a/european-commission-proposes-significant-reforms-to-gdpr-ai-act#:~:text=Board%2C,that%20mechanism%2C%20organizations%20would%20be
[2] iapp.org — https://iapp.org/news/a/european-commission-proposes-significant-reforms-to-gdpr-ai-act#:~:text=Board%2C,that%20mechanism%2C%20organizations%20would%20be
[3] 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

Key Statistics

- 42% of enterprises say over half their AI projects have been delayed, underperformed, or failed due to data readiness issues ([internalaudit360.com](https://internalaudit360.com/poor-data-readiness-is-plaguing-corporate-ai-projects/#:~:text=stuck%20managing%20infrastructure%20instead%20of,of%20enterprises%20say%20regulatory%20compliance)).
- 97% of organizations report active AI initiatives, but only 5% say their data is fully ready to support them ([www.cio.com](https://www.cio.com/article/4170978/nearly-every-enterprise-is-investing-in-ai-but-only-5-say-their-data-is-ready.html#:~:text=a%20new%20AI%20Momentum%20Survey,%E2%80%9D%20Early%20gains%20seen)).
- Poor data quality costs large enterprises an average of $12.9 million per year ([www.ibtimes.com](https://www.ibtimes.com/inside-data-foundation-problem-behind-enterprise-ai-failure-milan-parikh-takes-case-data-summit-3803265#:~:text=to,a%20new%20copy%20of%20the)).
- 60-70% of data teams run duplicate pipelines across departments, and bolting on governance later makes analytics 3-4x slower ([www.ibtimes.com](https://www.ibtimes.com/inside-data-foundation-problem-behind-enterprise-ai-failure-milan-parikh-takes-case-data-summit-3803265#:~:text=are%20difficult%20to%20dismiss,of%20that%2C%20you%20have%20already)).
- Bots account for 31% of all web traffic today, with roughly one-quarter of that being AI-related machine traffic ([techcrunch.com](https://techcrunch.com/2026/05/28/the-internet-is-being-rebuilt-for-machines/#:~:text=increasingly%20populated%20by%20agents,bot%20requests%20during%20that%20period)).
- Gartner predicts that by 2026, 60% of AI projects will be abandoned 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=Organizations%20that%20fail%20to%20realize,ready%20data)).

KEY TAKEAWAY

Amid rapid AI adoption, success depends less on models and more on data. Recent developments show that investing in modern data architecture, quality, and governance is now critical to scale AI and avoid costly failures or compliance setbacks.

Sources

The internet is being rebuilt for machines
https://techcrunch.com/2026/05/28/the-internet-is-being-rebuilt-for-machines/
Inside the Data Foundation Problem Behind Enterprise AI Failure: Milan Parikh Takes the Case to Data Summit 2026
https://www.ibtimes.com/inside-data-foundation-problem-behind-enterprise-ai-failure-milan-parikh-takes-case-data-summit-3803265
Nearly every enterprise is investing in AI, but only 5% say their data is ready
https://www.cio.com/article/4170978/nearly-every-enterprise-is-investing-in-ai-but-only-5-say-their-data-is-ready.html

Personal Agents Light the Fuse as Snowflake and Databricks Move Up the AI Stack
https://thecuberesearch.com/316-breaking-analysis-personal-agents-light-the-fuse-as-snowflake-and-databricks-move-up-the-ai-stack/
European Commission proposes significant reforms to GDPR, AI Act
https://iapp.org/news/a/european-commission-proposes-significant-reforms-to-gdpr-ai-act
Operational Data, Not Model Capability, Is the New Enterprise AI Moat
https://ai.via.news/enterprise-ai-infrastructure/operational-data-not-model-capability-is-the-new-enterprise-ai-moat
EU AI Act 2026 Updates: Compliance Requirements and Business Risks
https://www.legalnodes.com/article/eu-ai-act-2026-updates-compliance-requirements-and-business-risks
Poor Data Readiness Is Plaguing Corporate AI Projects
https://internalaudit360.com/poor-data-readiness-is-plaguing-corporate-ai-projects/
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