STATE OF DATA PRODUCTS

QUARTERLY RELEASE: COMMUNITY CURATION

Fixing the Fault-lines of GenAI

Why Data Architecture Matters More Than Ever

The urgency and direction to identify and solve the cracks underneath the surface of AI hype with potential data approaches and architectures..

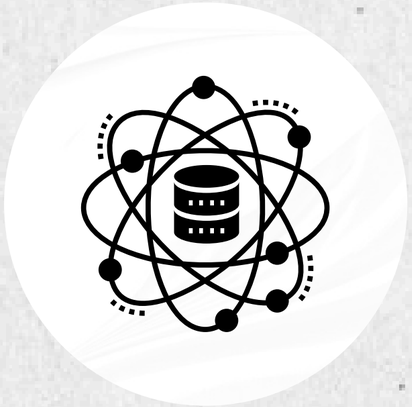
A TECH-REPORTING INITIATIVE
RITWIKA CHOWDHURY X MODERN DATA 101

WHAT'S IN STORE

The Challenges of GenAI

What VC's and Founders are Seeing

Deriving the Role of Data Solutions for GenAI



The Vision of Data Developer Platforms and Design Principles

"Operating Systems" for GenAI

...and so much more!

Editor's Note: 2025 Q2

Fixing the Broken Architecture of GenAI

It started with awe. When generative AI burst onto the scene, it was a tidal shift instead of just another wave of innovation. Boardrooms buzzed. Budgets ballooned. Pilots launched. Yes, so many pilots! And for a moment, it felt like anything was possible. The potential energy of AI at its wake was a historical feat.

But the euphoria didn't last. Somewhere between the 10th prototype chatbot and the 100th AI assistant demo, something broke. Not in the models, which are more powerful than ever. What broke was the connective tissue. The glue. The foundations of big, inflexible systems.

Today, enterprises find themselves caught in a paradox: GenAI adoption is rising, but its impact is stalling, slowing down, or most likely bringing in negative ROI due to huge hidden cost implications. Models constantly hallucinate, outputs have a good finish, but don't scale. Governance is running thin on all things AI, and ROI is elusive. What was promised as transformation now feels more like fragmentation.

Across industries, proof-of-concepts pile up but fail to scale. And business leaders are asking the question no one wants to answer: Why aren't we seeing the quick promised value that was supposed to get here as quickly as ChatGPT's conversational speed?

The problem, of course, isn't the "intelligence" of the models. These models are state-of-the-art frames that process like none before it. The problem was that these models were dropped into brittle architectures, ones never built to reason, adapt, or evolve. Data was treated as a raw resource, not a governed product that makes or breaks a model's outcome.

We are now witnessing the rise of systems like MemOS, where memory becomes governable, persistent, and foundational to agentic reasoning. The message is clear: if AI for enterprises is to deliver on its potential, enterprises must start building new systems with context, data, and continuity at the core.

From The Editor's Desk,
Modern Data 101

What 230+ Data Leaders & Practitioners Just Revealed

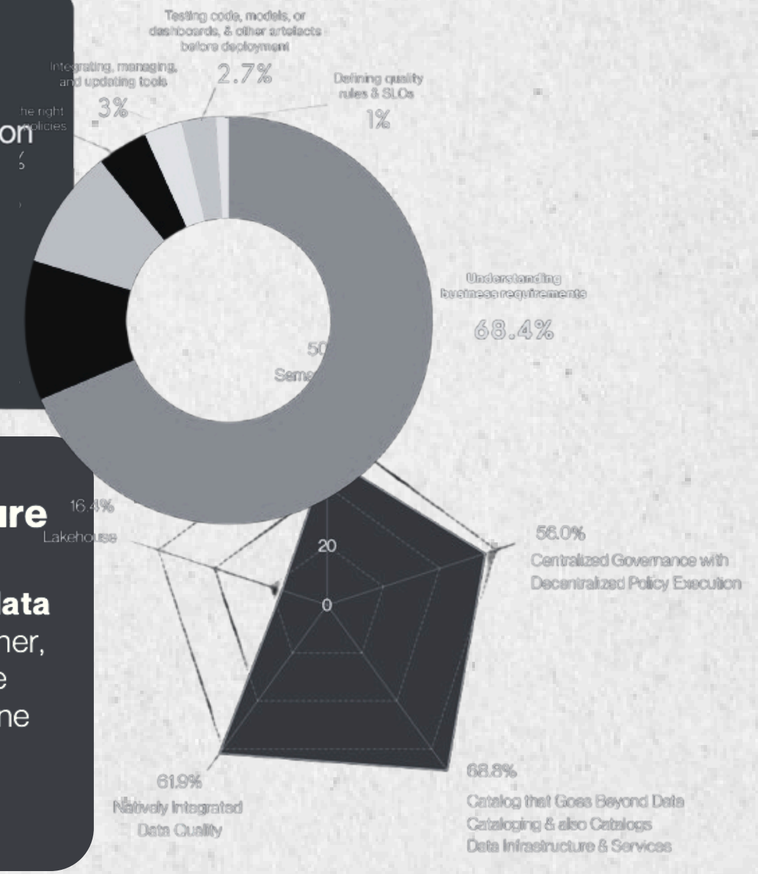
The data stack is no longer a growth engine; it's crumbling under its own weight. It's not a random opinion, but what 230+ practitioners across 48 countries, with an average of 15+ years of experience, shared. This is a collective reality check.

Conducted by the Modern Data 101 Community and facilitated by The Modern Data Company, The Modern Data Survey discovered that 70% of respondents said they want automation for stack orchestration. 85% say integrating tools across the stack is one of their top three challenges. That silent friction is delaying insights, increasing rework, and costing teams their competitive edge.

Access more insights in the survey report here↗️

Not Aspirations, Real Demands

- **70%** want automation for stack orchestration
- **60%** want integrated data quality
- **50%** need platform-native observability
- **70%** say cataloging beyond data, including infra + services is the next step forward.



Model-First, Product-Centric Future

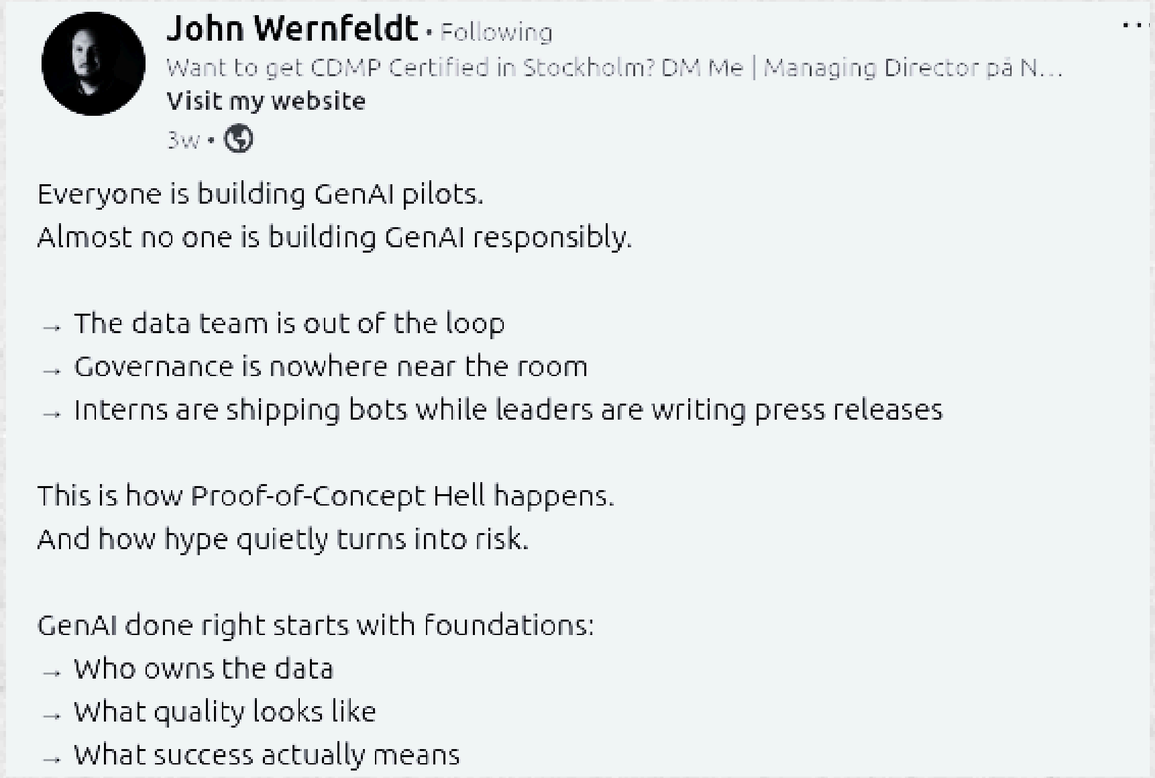
65% say the twin engines of acceleration are **data models and data products.** Not one or the other, but both working in unison. And to get there, we need to compress time, build trust, and reimagine the stack as a system of services: intelligently connected and business-aligned.

The GenAI Hangover: Lessons Learnt from Speed

In just under two years, GenAI went from buzzword to boardroom imperative. Startups announced funding rounds and shipped products in the same breath. Within weeks, five clones would appear: cheaper, open-source, already running ads. But that same momentum has become exhausting, even dangerous.

As Zain Jaffer, Founder of Vungle, noted, "Speed isn't an advantage anymore. It seems like a liability unless you have real lock-in." (Source)

What once looked like an arms race now feels more like a treadmill with no off switch. Inside enterprises, the speed has done more than sidelining caution; it is increasingly fracturing alignment. AI pilots are being built in silos, while data teams and governance functions remain out of the loop.



Source: John Wernfeldt on LinkedIn

Lessons from the 'Very Near Past'

When we talk about these challenges that industry veterans have faced, the PoC purgatory cannot go invisible. Without foundational involvement, these efforts drift. Technically clever but contextually lost. That context gap is showing up in a mounting wave of stalled Proof-of-Concepts (PoCs).

Founders mistake early interest for product-market fit. The result? A graveyard of demos that never scaled: what many now call "PoC purgatory." The issue isn't that PoCs are hard to land; they're everywhere. The real problem is what happens next: PoCs aren't converting into production. Teams are busy building GenAI demos, but those efforts are failing to scale, largely because data teams are out of the loop, governance is weak, and the models themselves lack guardrails and explainability.

Correcting this 'PoC Hell' in the words of John Wernfeldt, Managing Director - Data, Analytics and AI at Northridge Analytics, requires defining the right data ownership, its quality, and the correct success metric.

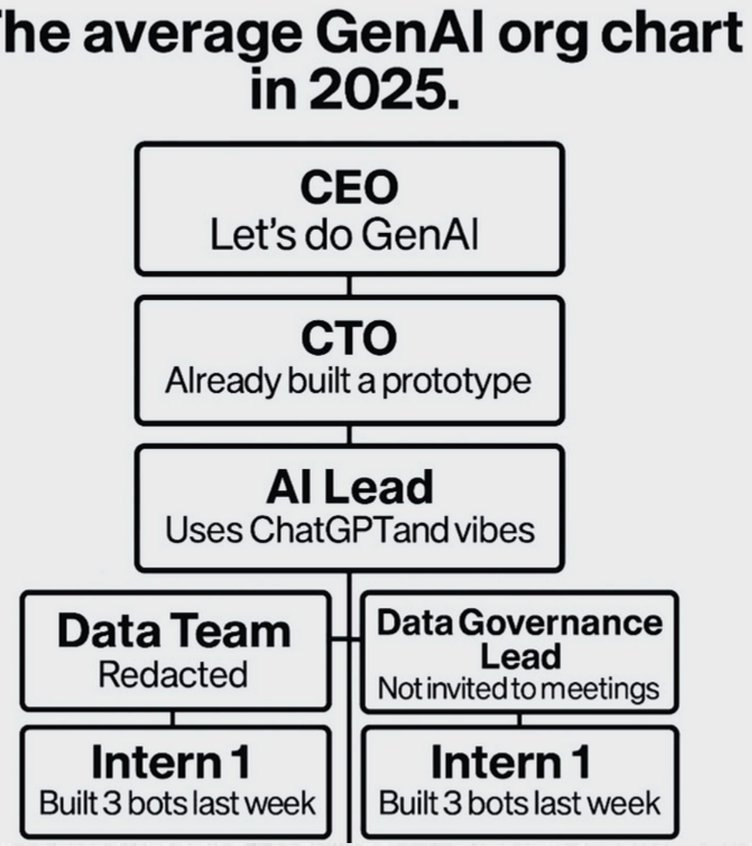
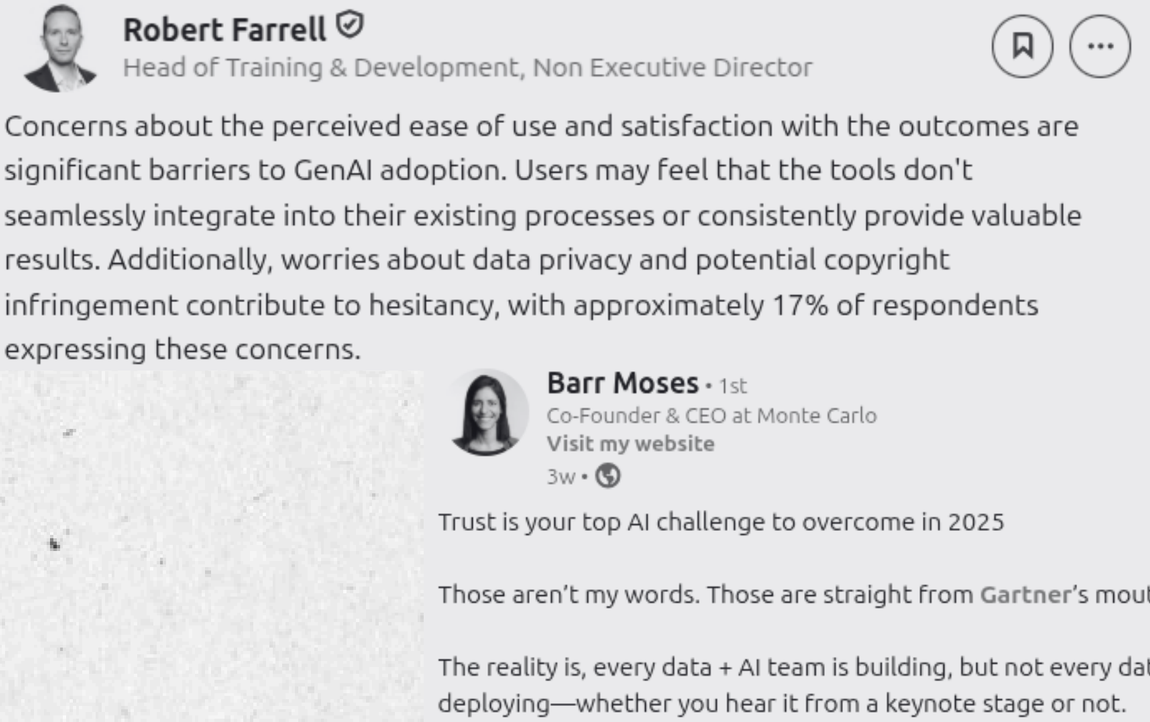


Image: The average GenAI org chart that assists an org to improve their AI strategy | Source

The Lessons: Are Customers Able to Trust AI?

Underpinning all of this is a growing trust crisis. Models may be smart, but they hallucinate. Outputs may be flashy, but they're often unverifiable. Organisations that once celebrated experimentation are now questioning if they can stake business-critical decisions on outputs they can't trace or explain.



Robert Farrell
Head of Training & Development, Non Executive Director

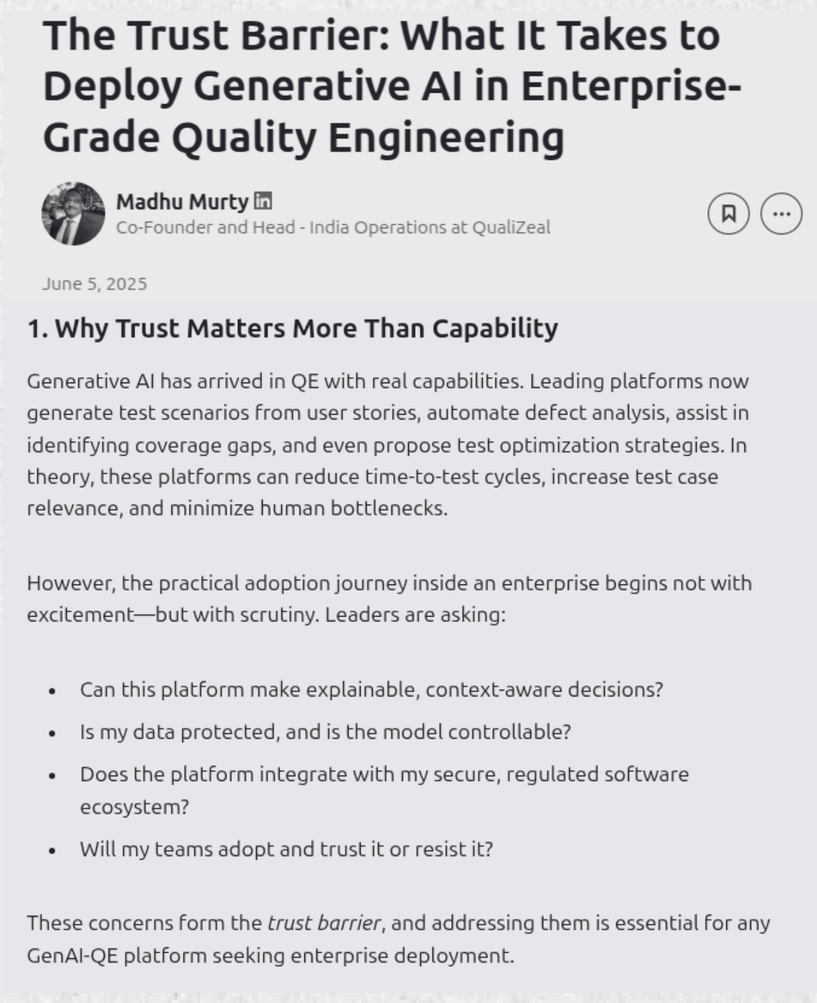
Concerns about the perceived ease of use and satisfaction with the outcomes are significant barriers to GenAI adoption. Users may feel that the tools don't seamlessly integrate into their existing processes or consistently provide valuable results. Additionally, worries about data privacy and potential copyright infringement contribute to hesitancy, with approximately 17% of respondents expressing these concerns.

Barr Moses • 1st
Co-Founder & CEO at Monte Carlo
Visit my website
3w •

Trust is your top AI challenge to overcome in 2025

Those aren't my words. Those are straight from Gartner's mouth (er, latest report?).

The reality is, every data + AI team is building, but not every data + AI team is deploying—whether you hear it from a keynote stage or not.



The Trust Barrier: What It Takes to Deploy Generative AI in Enterprise-Grade Quality Engineering

Madhu Murty
Co-Founder and Head - India Operations at QualiZeal

June 5, 2025

1. Why Trust Matters More Than Capability

Generative AI has arrived in QE with real capabilities. Leading platforms now generate test scenarios from user stories, automate defect analysis, assist in identifying coverage gaps, and even propose test optimization strategies. In theory, these platforms can reduce time-to-test cycles, increase test case relevance, and minimize human bottlenecks.

However, the practical adoption journey inside an enterprise begins not with excitement—but with scrutiny. Leaders are asking:

- Can this platform make explainable, context-aware decisions?
- Is my data protected, and is the model controllable?
- Does the platform integrate with my secure, regulated software ecosystem?
- Will my teams adopt and trust it or resist it?

These concerns form the *trust barrier*, and addressing them is essential for any GenAI-QE platform seeking enterprise deployment.

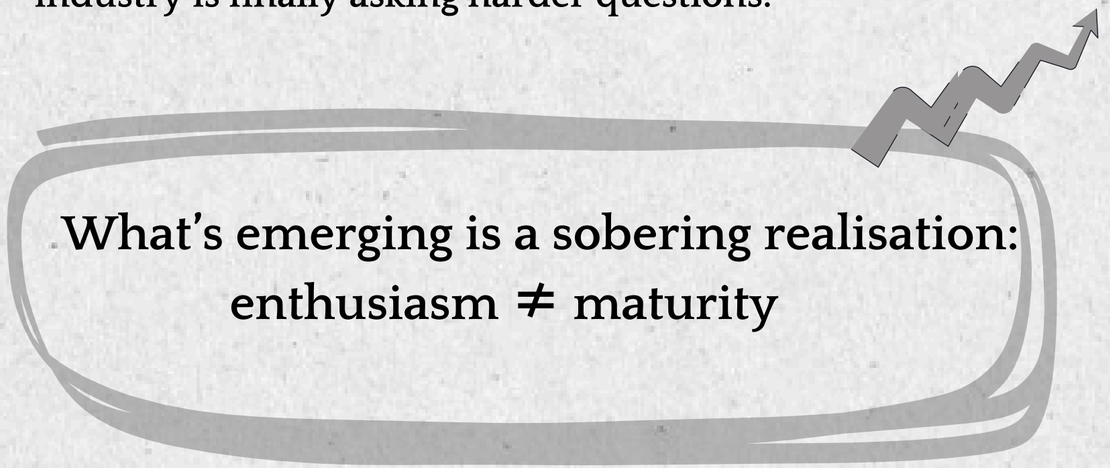
Too Much Enthusiasm, Not Much Maturity

The biggest irony? Adoption is accelerating, while readiness lags far behind. Popular reports name the tension plainly: maturity without measurement. GenAI is being hailed as strategic, but nearly half of orgs still don't track ROI. As the report warns, this is how transformation becomes just another buzzword: impressive on paper, underwhelming in practice.

"We're pouring money into models we can't explain, deploying tools we can't measure, and betting on Agents our orgs aren't ready to handle," mentions Stuart Winter-Tear, AI Strategy & Product Advisor at Dataception

We're moving beyond demos to platform-first thinking. That means shifting the emphasis from model capabilities to architecture, reuse, and value delivery.

The hangover isn't from AI itself; it's from the unchecked optimism that surrounded it. And as the dust settles, the industry is sobering up: the hype may still be loud, but the industry is finally asking harder questions.



VCs Waiting for Something Concrete. AI Wrappers are a No-Go.

Venture capital is no longer swayed by sheer potential. After a period of runaway enthusiasm, investors are pausing to ask harder questions: Where is the return? The hype-to-reality shift is clear. As David Prosser, Financial Journalist, noted in Forbes, companies are increasingly under pressure to prove ROI, not just promise it.

In fact, in a previously mentioned post by Zain Jeffer, Founder of Vungle, he also echoed his sentiment, observing that some VCs are choosing to wait, wary of flashy tools that don't hold up in enterprise settings. The earlier wave of startups, often wrappers around public APIs, delivered impressive demos but struggled with renewals, compliance, and scale. Speed, once seen as an edge, has become a liability when not anchored in defensible value.

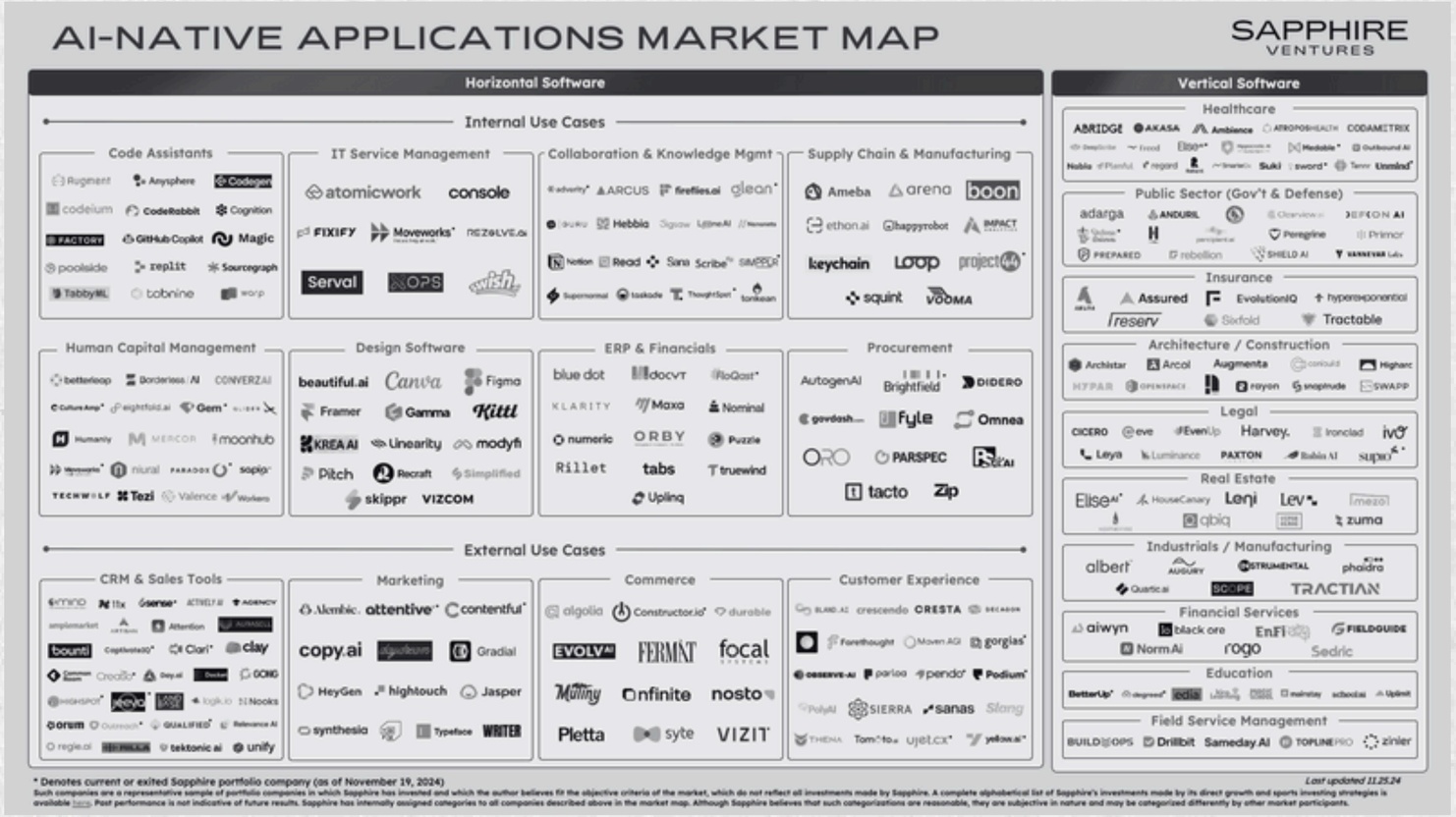


Image: A compilation of primary market opportunities of AI-native applications. (Source)

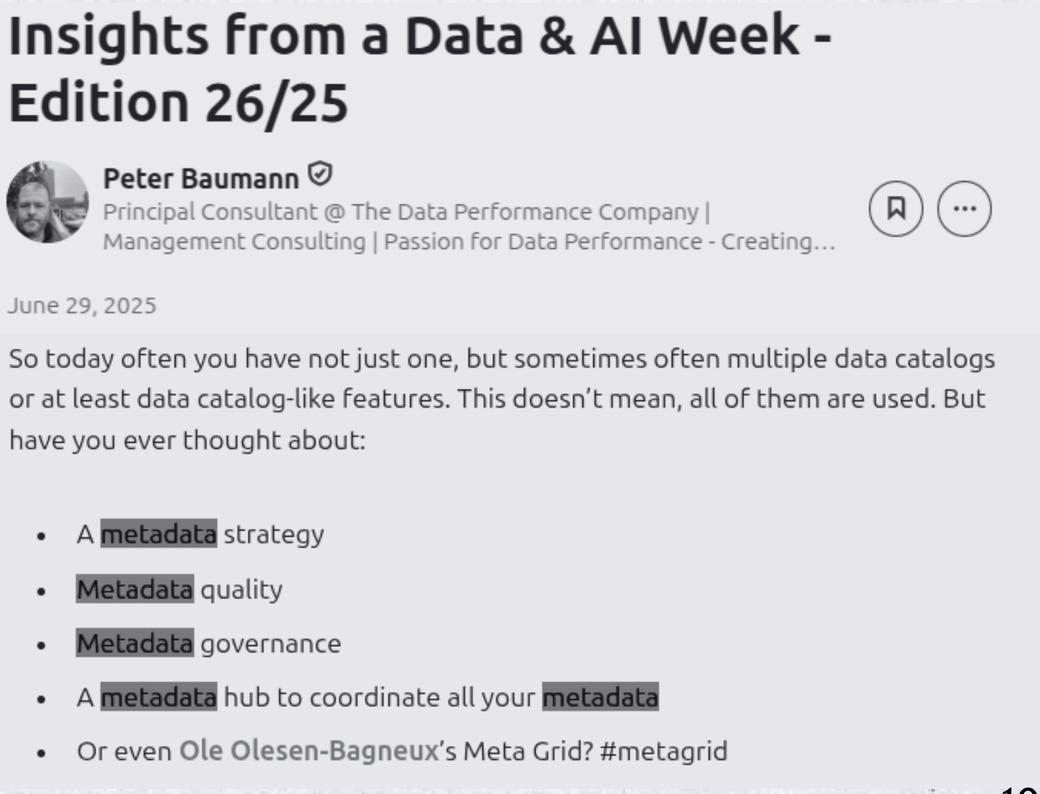
Tying Back GenAI Problems to Data Solutions

As organisations struggle with model hallucinations, lack of explainability, and governance blind spots, the root causes often trace back to weak metadata infrastructure. Now, the conversation is shifting: solving GenAI's biggest challenges means rebuilding trust through better data.
That's where metadata steps in, not just as documentation, but as the scaffolding for governance, reuse, and explainability. The problems that surfaced in GenAI are pushing organisations to finally invest in long-overdue data solutions.

In the early rush, lineage, policy enforcement, and data quality were often skipped in favour of speed. But as the costs of blind automation become visible, metadata is back at the centre, not as an afterthought, but as a design principle. Peter Baumann, aptly cautions,

"Don't take your metadata lightly. They are needed for AI, automation, business understanding and some other important tasks."

As organisations juggle multiple platforms: Databricks, SAP Datasphere, Denodo, Tableau, they're realising metadata sprawl without coordination leads to fragmentation, not trust.



Insights from a Data & AI Week - Edition 26/25

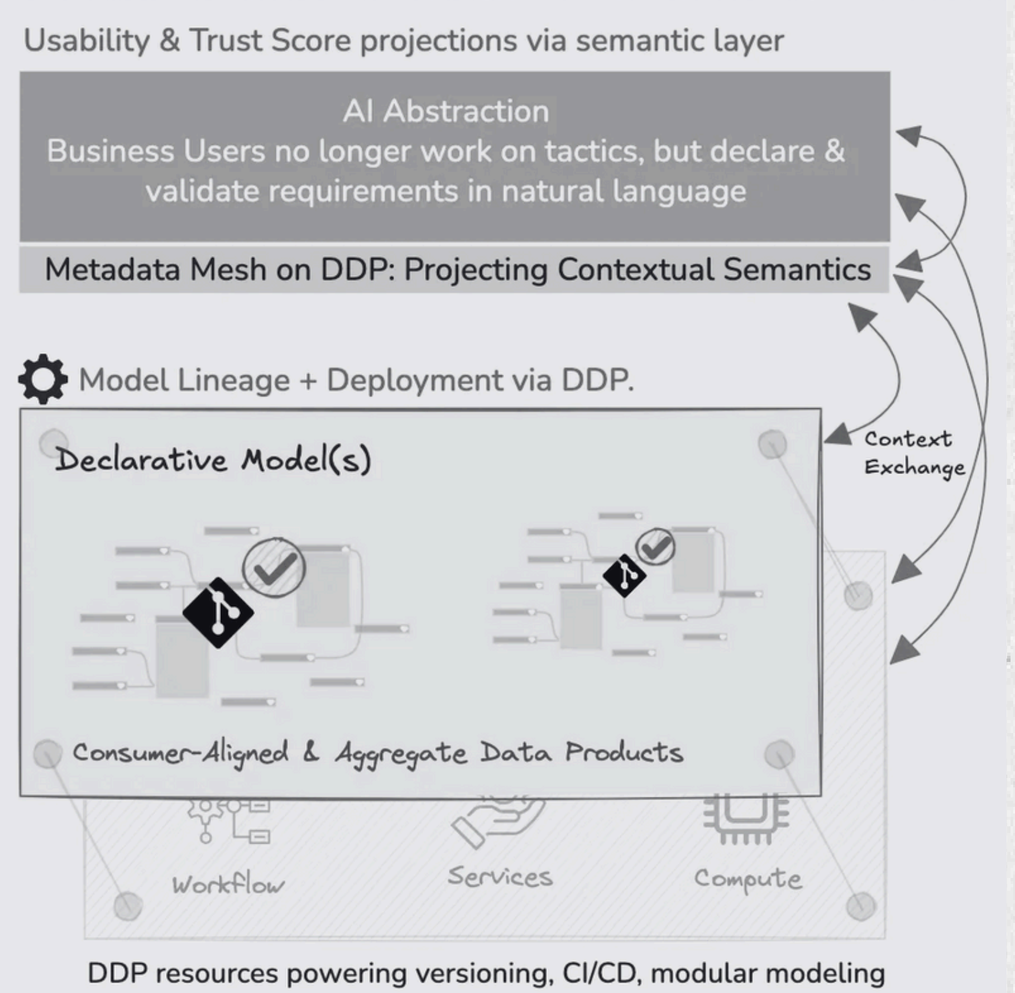
Peter Baumann
Principal Consultant @ The Data Performance Company | Management Consulting | Passion for Data Performance - Creating…

June 29, 2025

So today often you have not just one, but sometimes often multiple data catalogs or at least data catalog-like features. This doesn't mean, all of them are used. But have you ever thought about:

- A metadata strategy
- Metadata quality
- Metadata governance
- A metadata hub to coordinate all your metadata
- Or even Ole Olesen-Bagneux's Meta Grid? #metagrid

Thinking Metadata & Semantics for GenAI



DDP resources powering versioning, CI/CD, modular modeling

In the agentic era, metadata isn't just backend scaffolding, it's the connective tissue powering explainable, adaptive AI. As this section shows, data products and the semantic layer form a live, two-way loop, where data context (definitions, lineage, relationships) feeds AI systems, and AI interactions in turn generate new metadata, making quality, relevance, and trust emergent properties of the system.

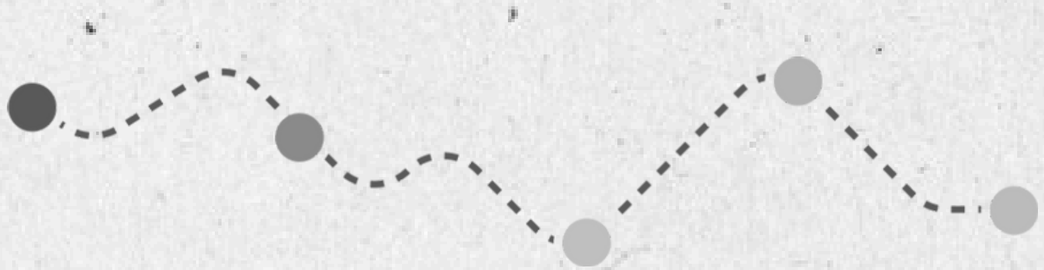
"User interactions generate new context and metadata, which flow back to data products and semantic models, sharpening their quality and relevance, mentions Animesh Kumar, CTO & Co-Founder at Modern.

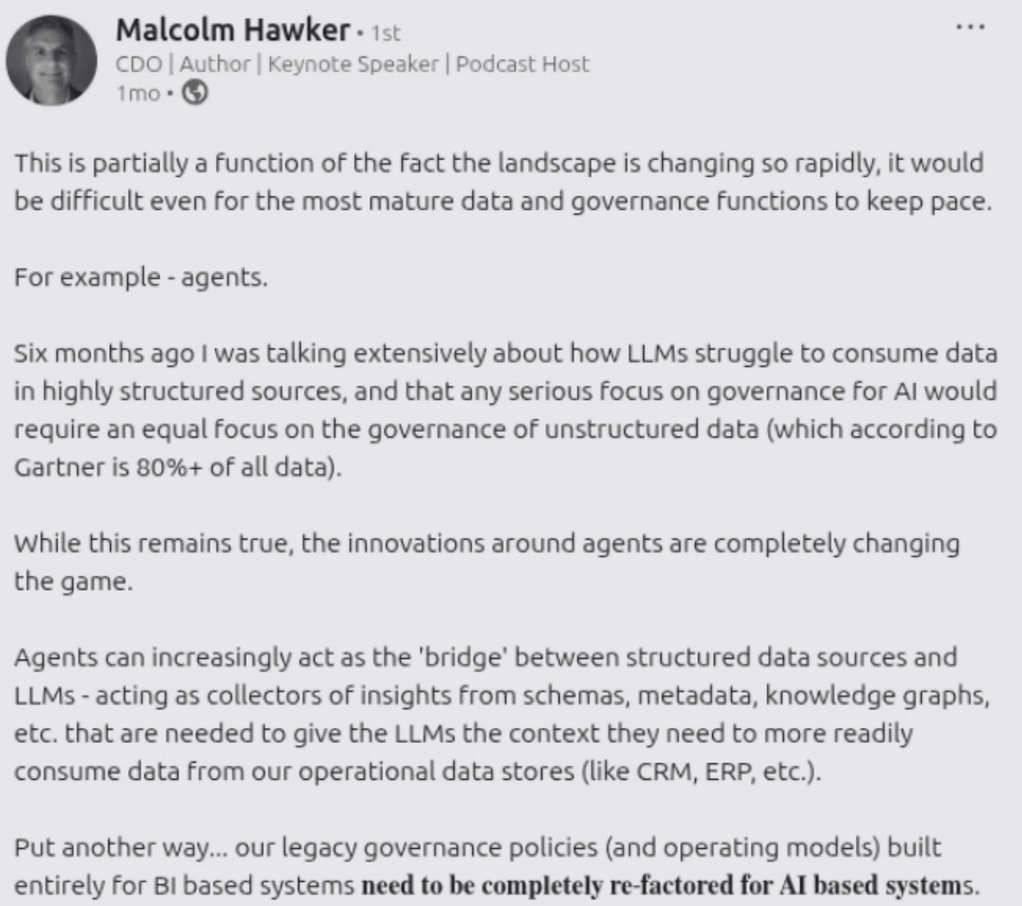
Image: A closer look (representation) of how the three forks interact for advanced and dynamic data quality | Source: Data Quality, a Cultural Device in the Age of AI-Driven Adoption

Does Metadata also Contribute to AI Governance

Metadata is becoming the new backbone of AI governance. It connects business domains to data producers, models to compliance requirements, and outputs to human validation loops. Those building future-proof AI aren't just investing in larger models; they're investing in semantic layers, contract-driven data flows, and participatory metadata practices. And perhaps most crucially, they're treating metadata not as a checkbox, but as a product in its own right, with clear ownership, lifecycle, and impact.

This evolution in metadata thinking also intersects with a bigger story: the redefinition of governance for AI-era systems. Most CDOs today aren't there yet, and that's okay. They're deliberately starting small, using limited-scope, internally-facing PoCs to contain risks while wrestling with emerging issues like explainability, policy automation, and even rethinking what quality means in AI contexts.





Malcolm Hawker • 1st
CDO | Author | Keynote Speaker | Podcast Host
1mo •

This is partially a function of the fact the landscape is changing so rapidly, it would be difficult even for the most mature data and governance functions to keep pace.

For example - agents.

Six months ago I was talking extensively about how LLMs struggle to consume data in highly structured sources, and that any serious focus on governance for AI would require an equal focus on the governance of unstructured data (which according to Gartner is 80%+ of all data).

While this remains true, the innovations around agents are completely changing the game.

Agents can increasingly act as the 'bridge' between structured data sources and LLMs - acting as collectors of insights from schemas, metadata, knowledge graphs, etc. that are needed to give the LLMs the context they need to more readily consume data from our operational data stores (like CRM, ERP, etc.).

Put another way... our legacy governance policies (and operating models) built entirely for BI based systems **need to be completely re-factored for AI based systems**.

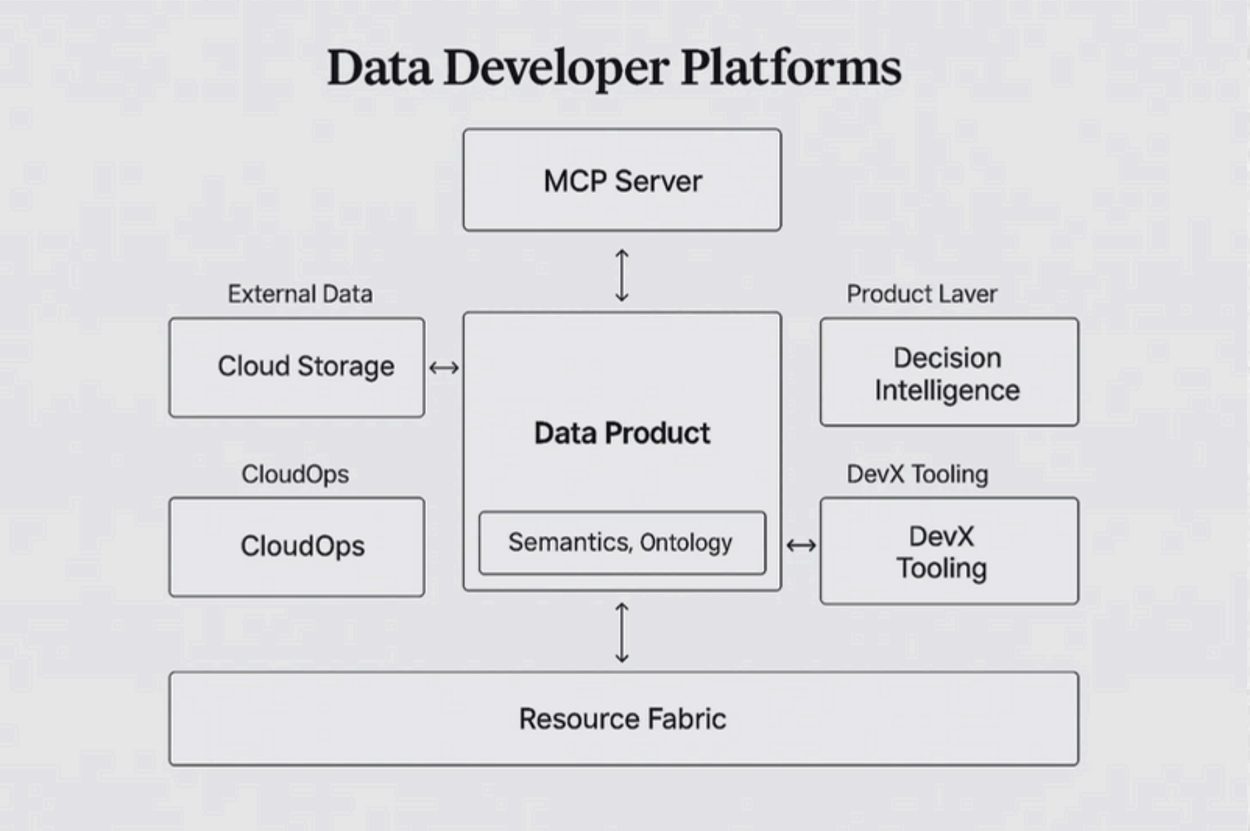
Pausing and Taking the U-Turn to Data as a Product

If we think of the case earlier, data delivery was driven by requests: “Build a dashboard,” “Create a report,” “Extract this table.”

In this approach, data work was reactive, siloed, and devoid of sustained business logic. But as GenAI introduces more autonomy and more risk, organisations are realising that task-based outputs no longer cut it.

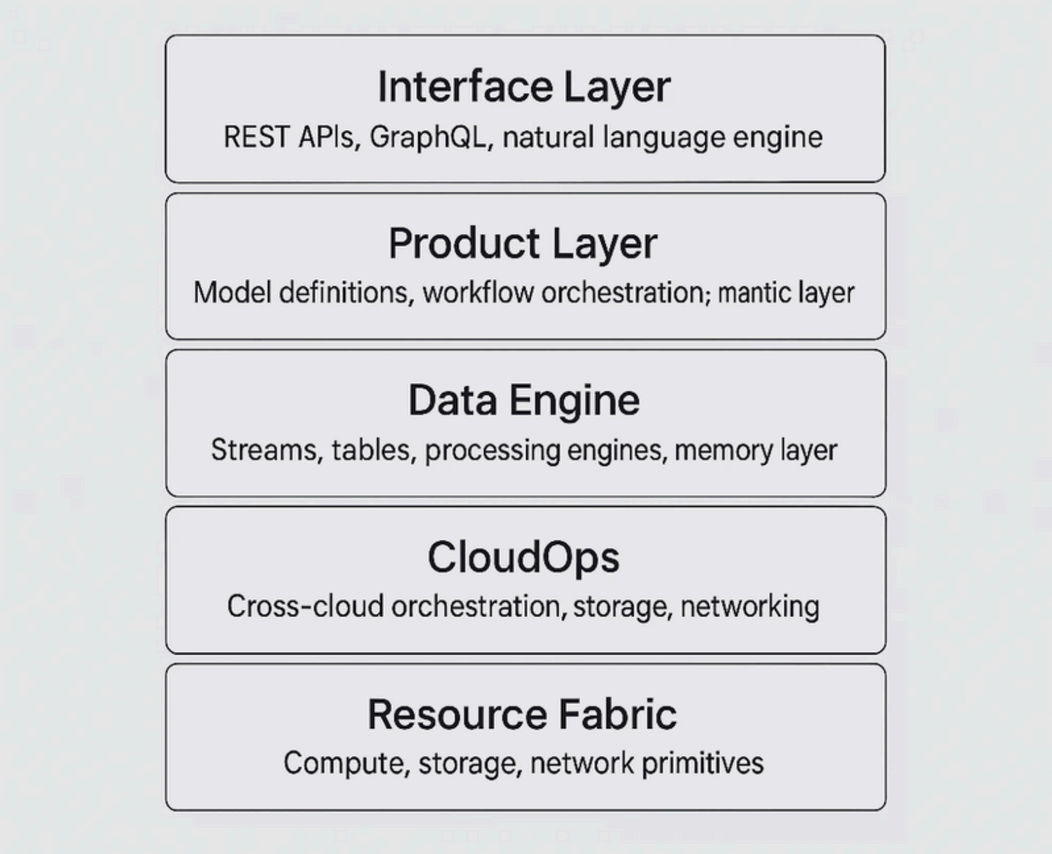
What’s emerging instead is a deliberate shift to productized data and AI, where every initiative, from a recommender to a chatbot, is treated as a long-lived product with measurable value, lifecycle ownership, and attached governance. Paolo Platter, Product Manager at Witboost, CTO - Co-founder at Agile Lab, notes this change:

“Data Products aren’t tasks. They’re investments.”



Overview of the DDP structure (for representation only) | Source: Thriving in the Agentic Era: A Case for the Data Developer Platform

Taking a U-Turn to Data as a Product Approach



Overview of the open layers of the Data Developer Platform architecture standard | Source: MD101 Archives

Amid the noise around models and multimodal breakthroughs, a quieter, foundational shift is underway: data is being rebuilt. As products with embedded intelligence instead of tables and dashboards.

Today's enterprise stacks weren't designed to reason. They describe the business, but they don't drive it. As a result, many GenAI initiatives are hitting structural walls—not for lack of ambition, but for lack of architecture. And that's where the data-as-a-product approach becomes not just relevant, but essential.

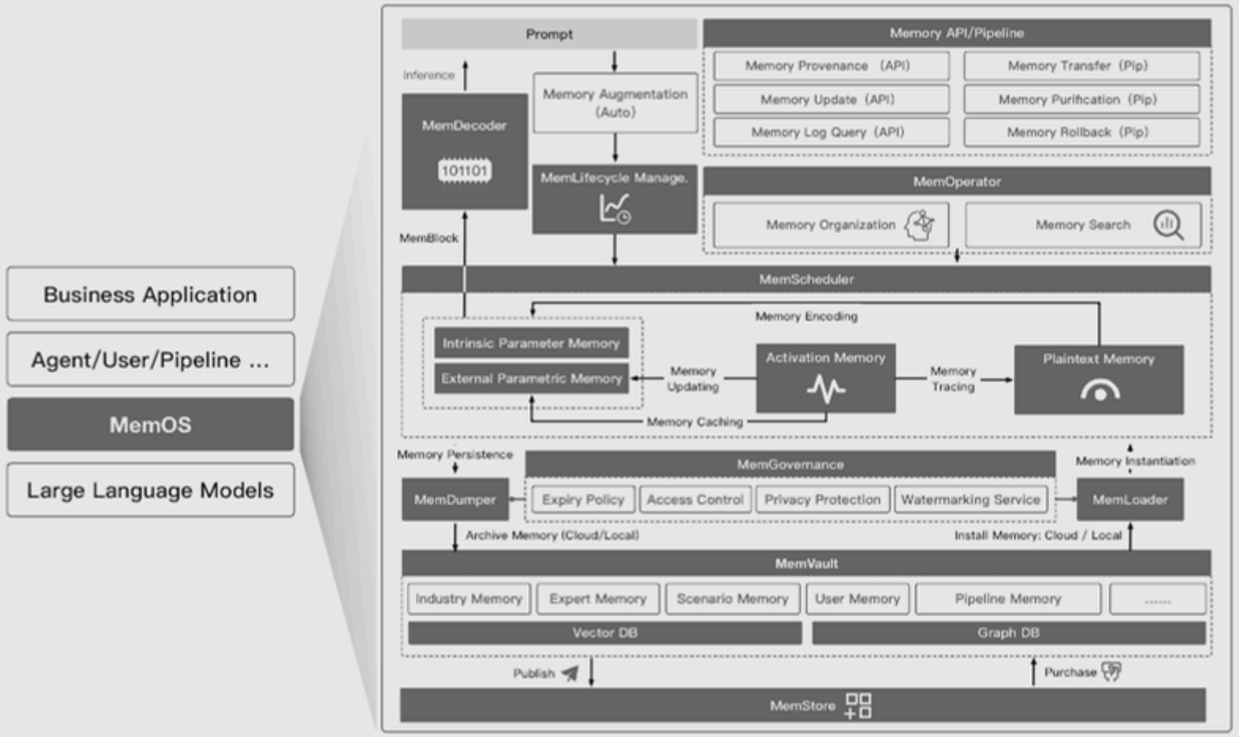
In a landscape where AI is expected to act, not just analyse, data products become more than pipelines; they become semantic assets. This means every product carries lineage, policy, usage context, and business meaning, all of which are prerequisites for responsible AI operations. These aren't static outputs. They are contract-backed, evolving, interoperable entities that AI agents can query, trust, and learn from.

Are you Hearing More about MemOS today? Here's why

An operating system like MemOS manages memory in LLMs through persistent, structured memory units, with version control, governance, and lifecycle management. It represents a vision where AI agents rely on structured, persistent, governable memory.

For such a memory architecture to function at enterprise scale, it cannot operate in isolation. What makes systems like MemOS significant is their emphasis on "agentic readiness": memories are not just logs but actionable, queryable, and updatable by agents through APIs. The system includes components for indexing, summarisation, retrieval, and consistency checking, all designed to support lifelong learning and reusability across tasks.

Where traditional vector DBs or cache-based memory struggle with context loss or incoherence, MemOS frames memory as a first-class OS-level abstraction, enabling agents to build internal models of the world that evolve. This structured memory layer is vital for advanced applications like autonomous customer support agents, legal copilots, or enterprise advisors that need to "remember" across sessions.



Overview of the MemOS architecture | Source: arxiv.org/pdf/2505.22101

The Idea of Operating Systems for Agentic AI Stacks

As we are witnessing an evolution in how platforms are imagined. Ideas like the Memory Operating System (MemOS) and Data Operating System (DataOS) point to a future where AI doesn't just consume data, it grows smarter by remembering, reasoning, and reusing knowledge across interactions and interfaces. One common context, multiple interfaces.

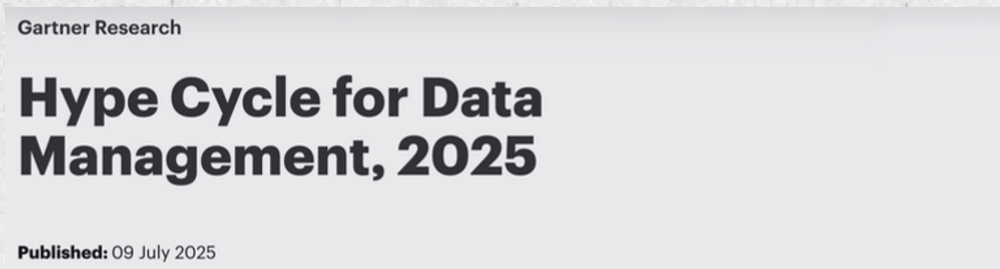
"Despite their remarkable capabilities in language perception and generation, current LLMs fundamentally lack a unified and structured architecture for handling memory." ~ Excerpt from paper on MemOS

Operating systems aren't plug-ins. They emerge from composable platforms where data acts like a product and memory functions as a utility. The real leap isn't faster output, but systems that retain, contextualize, and reuse knowledge like a living brain.

"...enabling autonomy for a host of personas is the number one objective of any operating system design...and one of many objectives can be established through autonomy across several layers of data producers and engineers." ~ Excerpt from philosophy on DataOS

This shift is what we should aim to capture; if the past few quarters were about acceleration, now it is about anchoring. The future of AI isn't built on faster models. It's built on data you can trust, explain, and evolve.

And this is no longer fringe theory. Gartner's Hype Cycle for Data Management, July 2025 (by Aaron Rosenbaum, Senior Research Director at Gartner, & Robert Thanaraj, Senior Director at Gartner), has formally recognised this directional shift, identifying a data operating system, DataOS by The Modern Data Company, as one of the leading solution providers anchoring this next wave of data architecture. The recommendations also includes Agile Lab, IBM, Informatica, K2view, Nexla, One Data, Revelate, & Starbust.



Top Stories on Modern Data 101 from Q2 2025

*Click tiles to read

Introducing Lakehouse 2.0: What Changes?

Transitioning to the 2nd Generation, an Architectural Overview, Pivots for Decentralisation, Design that brings...

APR 18 · ANIMESH KUMAR AND TRAVIS THOMPSON

49

Data Lineage is Strategy: Beyond Observability and Debugging

Gaps in passive lineage, how and why Data Products change and uplift lineage, and notes on stepping up to the AI-native...

MAY 1 · TRAVIS THOMPSON

41

The Role of the Data Architect in AI Enablement

Pivoting on right-sized foundations, delivered through focused vertical slices, and driven by impact, not just...

MAY 22 · COLIN HARDIE

31

Data Quality: A Cultural Device in the Age of AI-Driven Adoption

Transitioning to proactive quality with a trifecta of AI abstractions, productised data tracks, and self-serve...

MAY 30 · ANIMESH KUMAR

32

Universal Truths of How Data Responsibilities Work Across Organisations

Building higher clarity on data accountability, observation of common ownership patterns across large data organisation...

JUN 5 · PAUL JONES

29 1

Join The Modern Data 101 Crusade!

If this playbook was helpful for your implementation or even as a means to organise your thoughts, we'd highly recommend you join us and help us build the Data Product Expertise in the data industry.

We are ALL THINGS DATA PRODUCTS!

Find a truly community-driven voice with experts from all walks of the data industry voicing their insights in the Data Product space on our channels. We'd love for you to come forward and share your insights with us so we can improve our delivery and truly create for YOU.

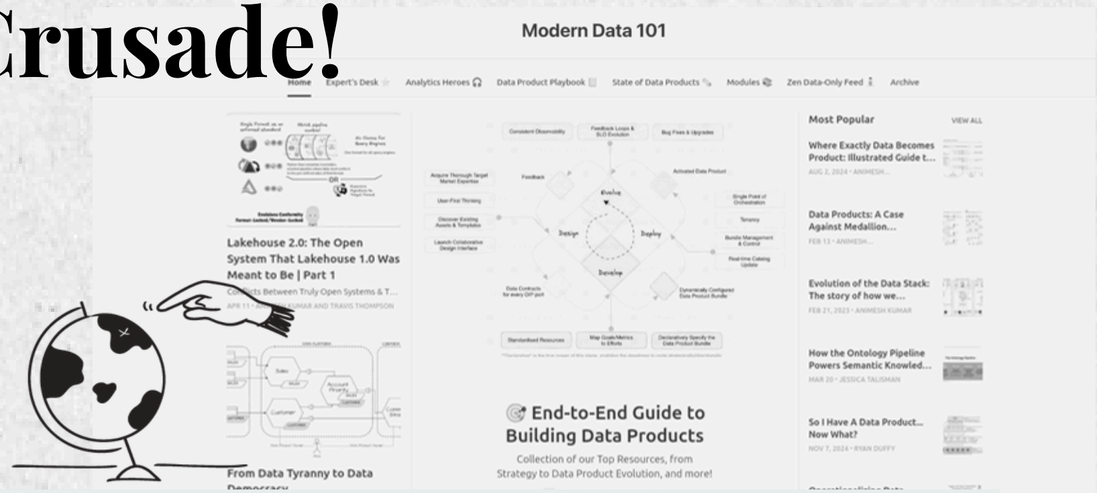
Subscribe

9000+ Data Peers

100+ Resources

30+ Contributors

120+ Countries Reached



Your Community, Your Benefits

Explore a world of reliable information, connect with experienced peers, share insights, boost visibility, receive valuable feedback, and propel your career forward in the ever-evolving data industry.

Access To Resources

In an age of infinite data perspectives, have trustworthy information to make a judgment.

Opportunity to Network

When learning about data, surround yourself with people more experienced than you.

Data Product Expertise

Find all things data products, be it strategy, implementation, or a directory of top data product experts & their insights to learn from.

Visibility & Recognition

Showcase your expertise, contribute to discussions, and build your reputation within the community.

Career Growth

Stay updated on the latest trends and job opportunities in the data industry, and gain valuable career advice.