NEW EDITION

THE CATALYST

WHERE DATA & AI IGNITE BUSINESS

STATE OF DATA PRODUCTS

A Tech-Reporting Initiative

Ritwika Chowdhury x Modern Data 101

SPECIAL EDITION

JANUARY 2026

Founder's Note

THE CATALYST

WHERE DATA & AI IGNITE BUSINESS

This issue is packed with the innovations and newly emerged concepts that disrupted the data & AI landscape in 2025.

I write to you as a fellow witness to a year of quiet but consequential change in data and AI. The State of Data Products 2025 goes beyond a retrospective and distils how evolution is striking us at a record pace. We can confidently say that technology is at its fastest rate of growth than it ever has in history.

The clearest insight of 2025 for us has been that data products have crossed an invisible threshold. The conversation has shifted decisively from how to build them to how they behave in the real world. Teams are no longer asking for more tools, but more coherence. Governance is being pulled closer to product thinking, metadata is being reinterpreted as user experience, and AI has turned data quality from a back-office concern into a front-line business risk.

The strongest data products this year weren't the most sophisticated, but the most legible, observable, and resilient under change. And on a more personal note, what stood out to me most was how dramatically user expectations accelerated.

Consumers of data, human and machine, now assume immediacy, adaptability, and context by default. This edition is shaped by that realisation: that the future of data will belong not to those who optimise pipelines fastest, but to those who design for context and, most importantly, understand how AI-readiness is a whole different ball game than Analytics-readiness.

Animesh Kumar

Journey of Data Products and AI through 2025

Q1

IDENTIFYING AI READINESS

- Lack of data readiness
- Limitations of data stacks and platforms
- Rise of new architectural designs like Lakehouse 2.0

Q2

ARCHITECTING NEW DESIGN PRINCIPLES

- Fragmented tooling and stack environment
- Changing VC and investor trends
- Increased attention towards semantics
- Aligning operating systems like MemOS and DataOS for AI enablement

Q3

RE-PREPARING FOR AI

- Failure points of data platforms
- New platform paradigms and modelling approaches for AI
- Popularity of context engineering

Q4

DATA BECOMES THE CONTROL PLANE

- Governance debt operationalised
- Data elevated to business strategy
- Realigning data products
- Marketplaces became default

2025
KEY PIVOTS

Rise of Agentic AI

New Architectural Design

Era of Context Engineering

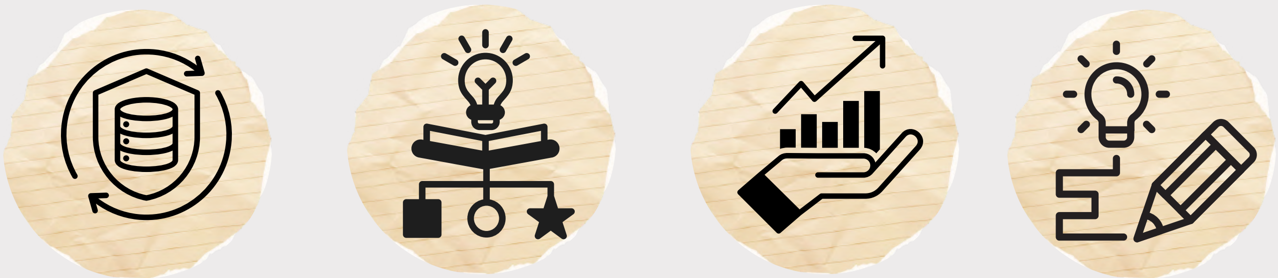
New Challenges of Data Platforms

Addressing New Governance Gaps

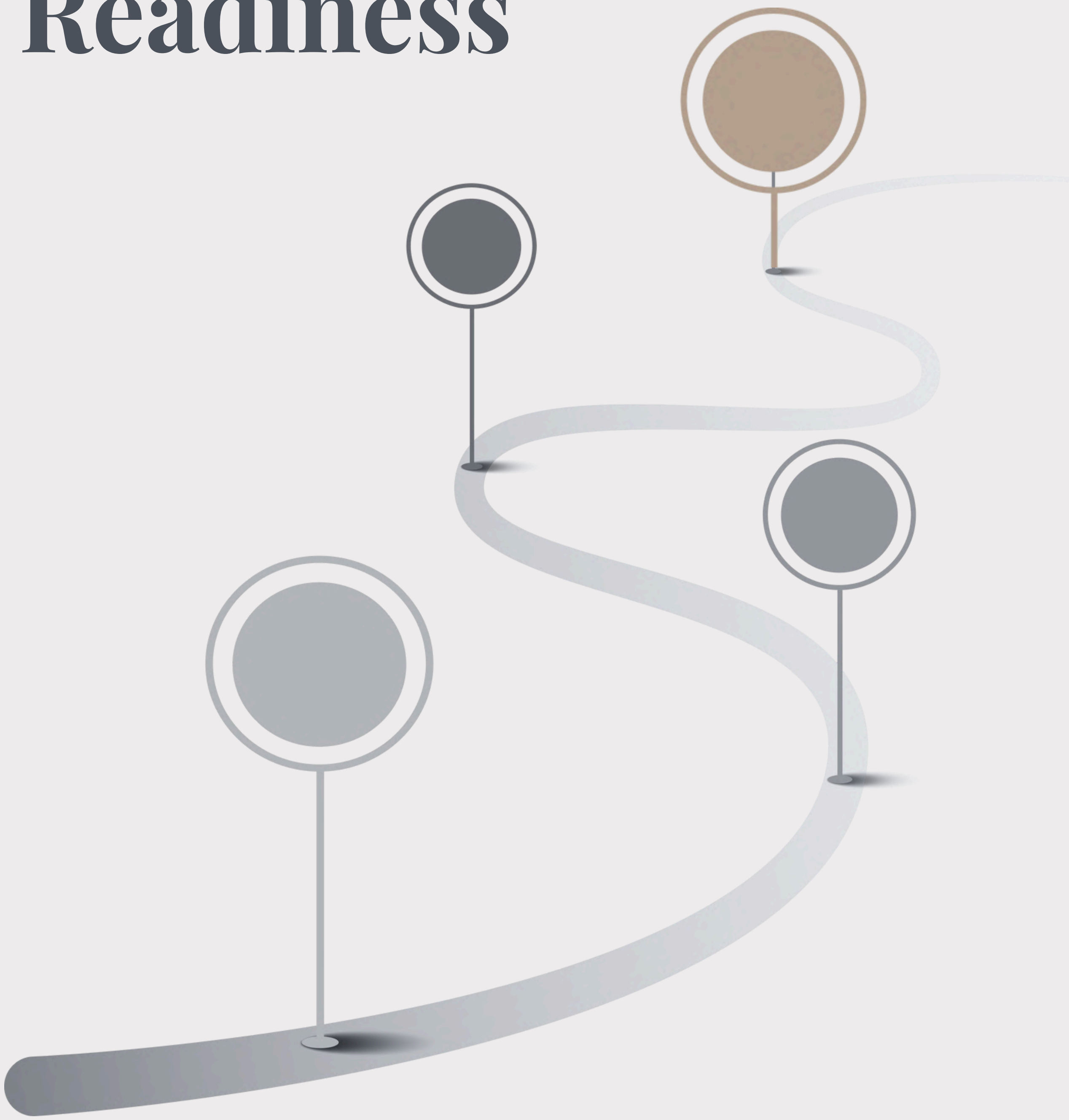
Increased Focus on Semantics

Data Products as Business Strategy

Redesigning Data Movement



Q1
Identifying AI Readiness



Agentic AI in Vogue

Agentic AI witnessed a massive popularity, indeed, much more than expected! Tech enthusiasts, engineers, decision-makers, and top leadership all showed interest and walked straight into the space.

According to Accenture, by 2030, AI agents will become the "primary users" of most enterprise systems effectively acting as digital colleagues, interfacing with software on humans' behalf (Source).

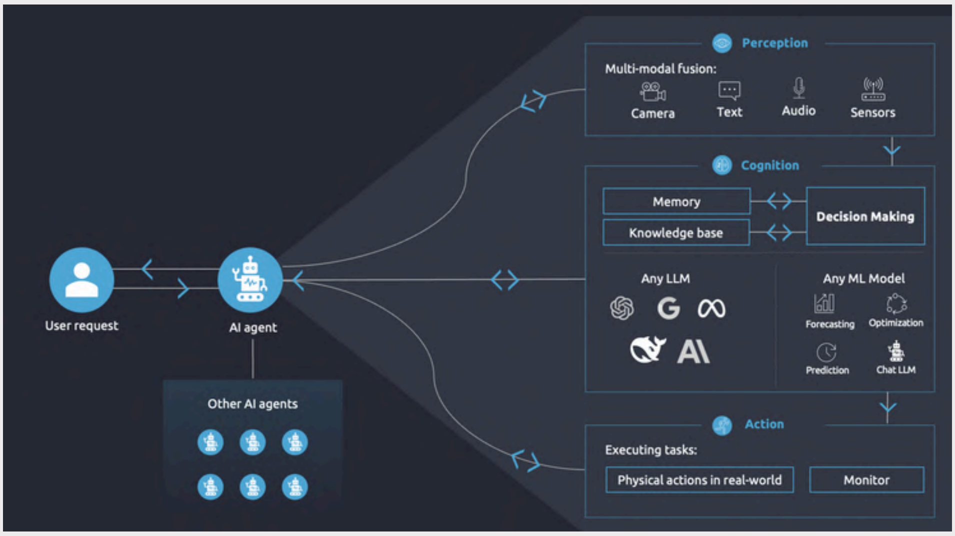


Image: Simplified Overview of AI Agent Design | Source: Bora Ger, Global AI Accelerator Upskilling Lead, Capgemini Invent & Olexander Tokunov, Senior Manager at EY.

But several questions started emerging, such as "How would enterprises retain that personal touch if frontline support is AI," which is a true concern even today for many. Could these agents actually pick up the feeling of real humans? If yes, that's a fear too!

However, amidst all the fascination for agents and GenAI, what actually happened was, **Agentic AI entered enterprise conversations before the foundations were ready.**

Most organisations encountered agents as Auto-GPT-style workflows stitched together with scripts, or "AI coworkers" positioned as accelerators, and not accountable systems.

This did create excitement, but without containment.

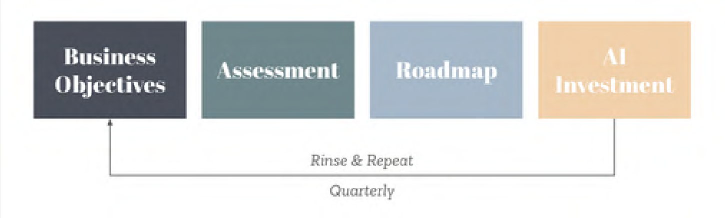
But as technological change accelerates with transformative forces like AI, there should be real fear that the risk isn't just falling slightly behind; it's being left permanently stranded," mentions Will Hansmann, CTO at Upshop

In this context, every business leader needs to assess their readiness. The critical insight executives can't afford to ignore."

The Readiness Gap That Defined 2025

The biggest failure of AI readiness in 2025 wasn't ambition, tooling, or talent. It was the absence of governance that could keep pace with experimentation. Standardised procurement, integration, security testing, and lifecycle ownership lagged far behind adoption velocity. As decentralised teams raced to deploy models and agents, enterprises quietly accumulated ungoverned risk.

The second readiness failure was structural: data existed, but semantics didn't. Organisations assumed models would compensate for inconsistency, fragmentation, and misalignment. Instead, AI systems magnified these flaws. Agentic use cases failed not because of intelligence limits, but because shared meaning, context, and standards were missing.



Source: Stephen Sklarew

"By the time AI implementation begins, your costly strategy documents may already be out-of-date," mentions **Stephen Sklarew**, **Chief Executive Officer** at Synaptiq.

Adoption outpaced actionability!
Despite rapid growth in AI usage, very few organisations reached meaningful scale. Skills gaps, poor data quality, and brittle integrations stalled progress. The result: high activity, low impact. Enterprises could deploy AI, but struggled to operationalise it reliably.

The signal was clear: **readiness without actionability is not readiness at all.**

"Too many leaders are still 'waiting and seeing' on AI. Wake-up call: the water's pulled back, and the wave is barreling in," mentions Will Hansmann, Chief Technology Officer at Upshop.

In 2025, part of the AI-readiness problem wasn't capability, but urgency. Many organisations remained in "wait and see" mode even as AI moved from novelty to operational force.

Early adopters are not just experimenting, but were realigning stacks, breaking down data silos, and building for agentic workflows right now. The competitive divide is widening between those prepared to act and those still hesitating.

The Readiness Gap That Defined 2025

Laurent Dresse · 1st
The Data Governance Kitchen | Chief Evangelist | Thought Leader | Da...
Visit my website
10mo • Edited •

Can your data truly be considered AI-ready?

In the rapidly evolving landscape of artificial intelligence, having AI-ready data is more than just a buzzword—it's a necessity. Organizations are racing to harness the power of AI to drive insights and innovation, but many stumble at the foundational stage: data readiness. Ensuring your data is clean, comprehensive, and compliant sets the groundwork for successful AI implementations. It's about creating a robust ecosystem where data can flow seamlessly through all stages of processing and decision-making.

Joel Bearden · 3rd+ 10mo
Sr. Director, Technology | Product Security at Cox Communications

Nah - the biggest risk is adopting AI without understanding insecure design or implementation. Change my mind. 😉

AI is a force-multiplier for the business; without guard-rails, the same applies for threat agents.

Mike Fishbein · 3rd+ 10mo
Building custom AI agents for marketing and sales teams

The AI skill gap is interesting. It seems like the best way to learn in these early days is by jumping in and doing it and chatting with others are doing the same. There aren't that many great resources, and everything changes so fast because of all the new launches.

Tiankai Feng · 1st
Data & AI Strategy Director @ Thoughtworks | Author of "Humanizing ...
9mo • Edited •

My main message was that we can't just drive AI through excitement for the technology, but we need to consider the "bigger picture" - business alignment, workforce literacy, operating model, architectural foundations and proactive governance - to truly drive AI through collective purpose.

Fragmented Data Stacks as a Structural Bottleneck

The 1st quarter of 2025 witnessed how tool sprawl and vendor fragmentation increase operational drag, erode trust through inconsistent semantics and governance, and slow decision-making, ultimately capping data ROI and limiting AI readiness.

When data flows across disconnected tools, semantics fracture, governance becomes inconsistent, and lineage loses continuity. **AI systems inherit this fragmentation**, producing outputs that are technically correct but operationally unreliable.

Tooling tooling sprawl did come up as a cost and complexity issue. With passing time, the industry reframed it as a systems design failure. Adding more tools didn't increase capability but increased coordination overhead.

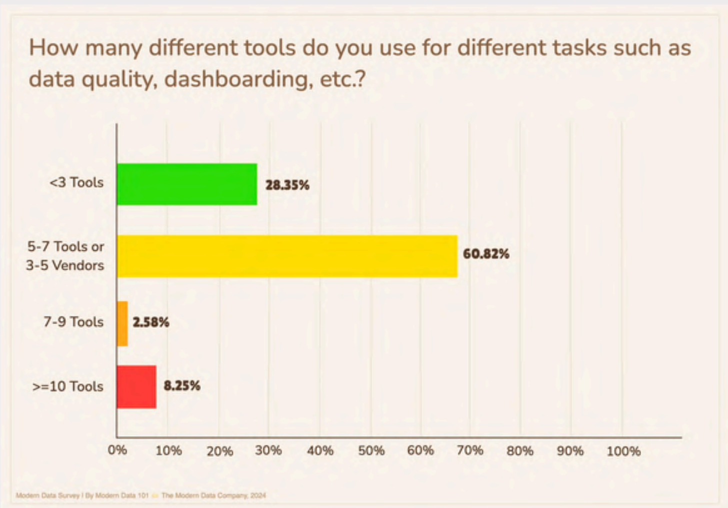
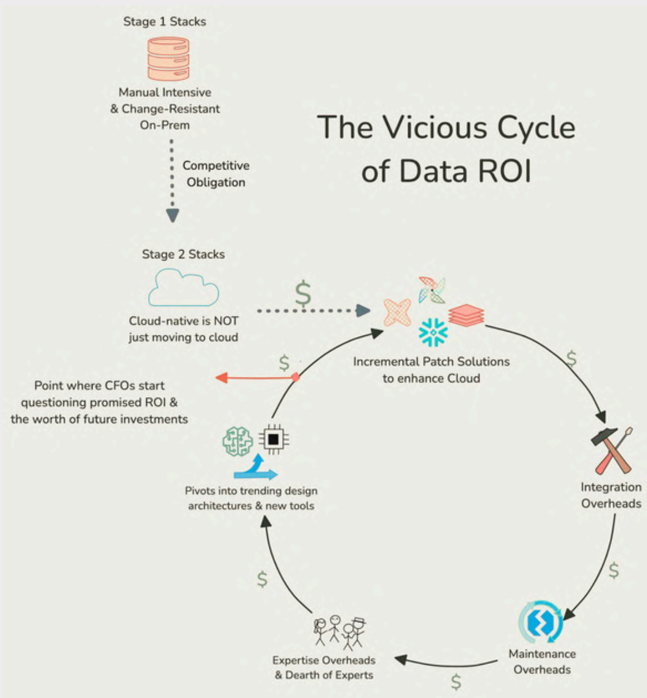


Image: The Modern Data Survey, 2024-25

The result was slower experimentation, brittle integrations, and stalled AI initiatives.



The image shows the vicious cycle of diminishing data ROI, a consequence of fragmented or siloed tooling and data architecture | Source: Srujan Akula

Redirecting Lakehouse Architecture as AI-Enabler

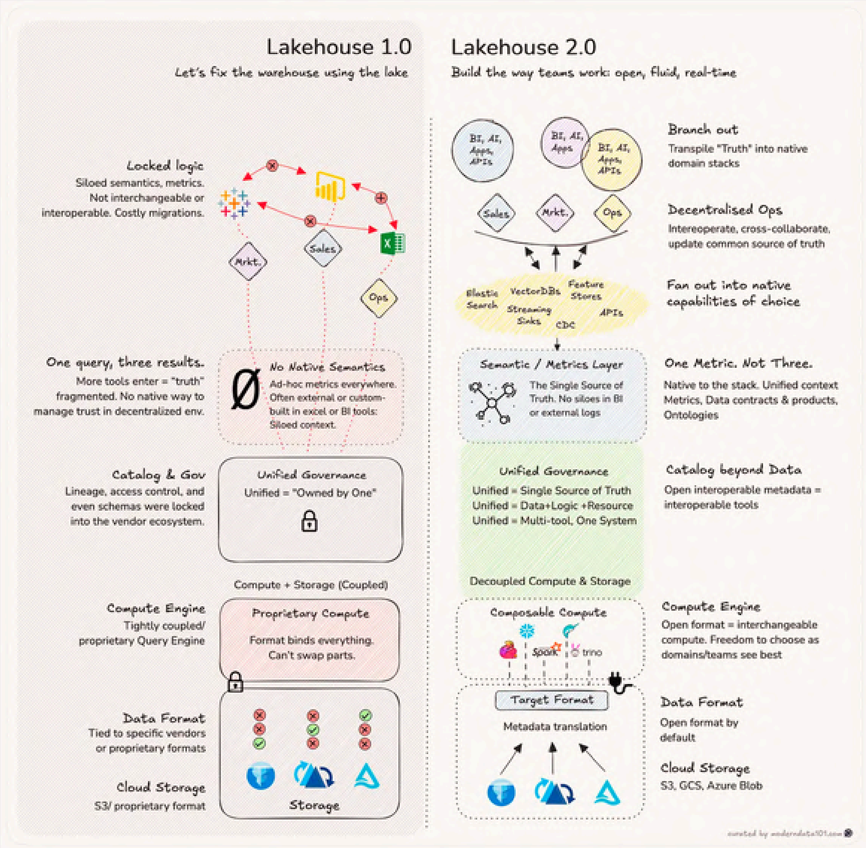


Image: Lakehouse 2.0 vs. Lakehouse 1.0 | Source: Introducing Lakehouse 2.0: What Changes

Some industry leaders made an early prediction of how conventional architectures report challenges with legacy, monolithic architectures and advocate for modern design principles. A critical problem that surfaced was the limited scope of Lakehouse 1.0.

"In the Lakehouse 1.0 Vision:

✅ Everything follows strict rules.

✅ Locked into one format, ensuring consistency.

✅ The system is optimised, but only for those willing to conform." - Animesh Kumar in The Open System Lakehouse 1.0 Was Meant to Be

This required modularity, self-service, and composability, marking a philosophical shift from static pipelines to purpose-driven data systems. The result was slower experimentation, brittle integrations, and stalled AI initiatives.

As AI-readiness became a priority, organisations realised that conventional lakehouses were not built for adaptability. While they promised unified, lower-cost architectures by combining lakes and warehouses, reality exposed structural limits.

Vendor lock-in restricted engine choice across compute, storage, and catalogs. Storage–compute coupling forced query engines to conform to storage logic. Table formats were tightly bound to specific engines, limiting interoperability and modularity.

These constraints made traditional lakehouses rigid, difficult to evolve, and poorly suited for AI-driven workloads that require flexibility, openness, and rapid change.

Lakehouse 2.0: A Modular, Open Architecture for AI

Lakehouse 2.0 represents a shift toward modular, open, and interoperable data architectures designed for AI-era demands. Instead of monolithic platforms, it promotes composable, best-of-breed stacks where storage, compute, governance, and query engines are decoupled.

“The urgency is different, but just as real. This time, it’s the AI race driving acceleration. Every company needs (more than “wants”) to deliver faster insights, smarter applications, and real-time intelligence. And the old constraints are too heavy to carry forward,” mentions Animesh.

By embracing open table formats such as Apache Iceberg, Hudi, and Delta (as open specifications), Lakehouse 2.0 enables ACID transactions, schema evolution, and consistent metadata handling while allowing multiple engines like Spark, Trino, DuckDB, Snowflake, and others to operate on the same data without pipeline rewrites.

This decoupling restores freedom of choice and reduces long-term cost and lock-in. Teams can plug and replace engines, catalogs, or governance layers

Capability	Lakehouse 1.0	Lakehouse 2.0
Domain Collaboration	Disconnected & ad-hoc	Federated & semantic
Semantic Layer	Absent or locked in BI	Native and open
Compute Choice	Vendor-locked	Composable & domain-aligned
Catalog	Siloed & centralized	Federated & interoperable
Tooling	Locked vertical stack	Fan-out to best-of-breed tools

Image: Comparative analysis of Lakehouse 1.0 vs. 2.0

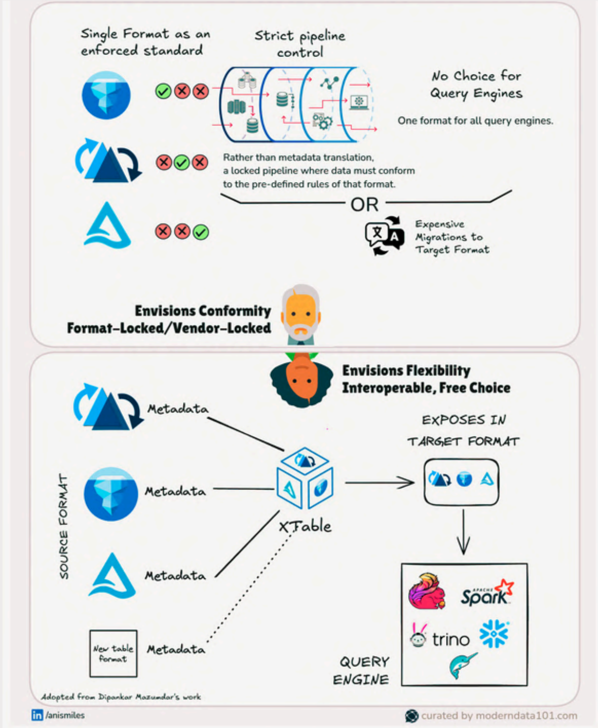


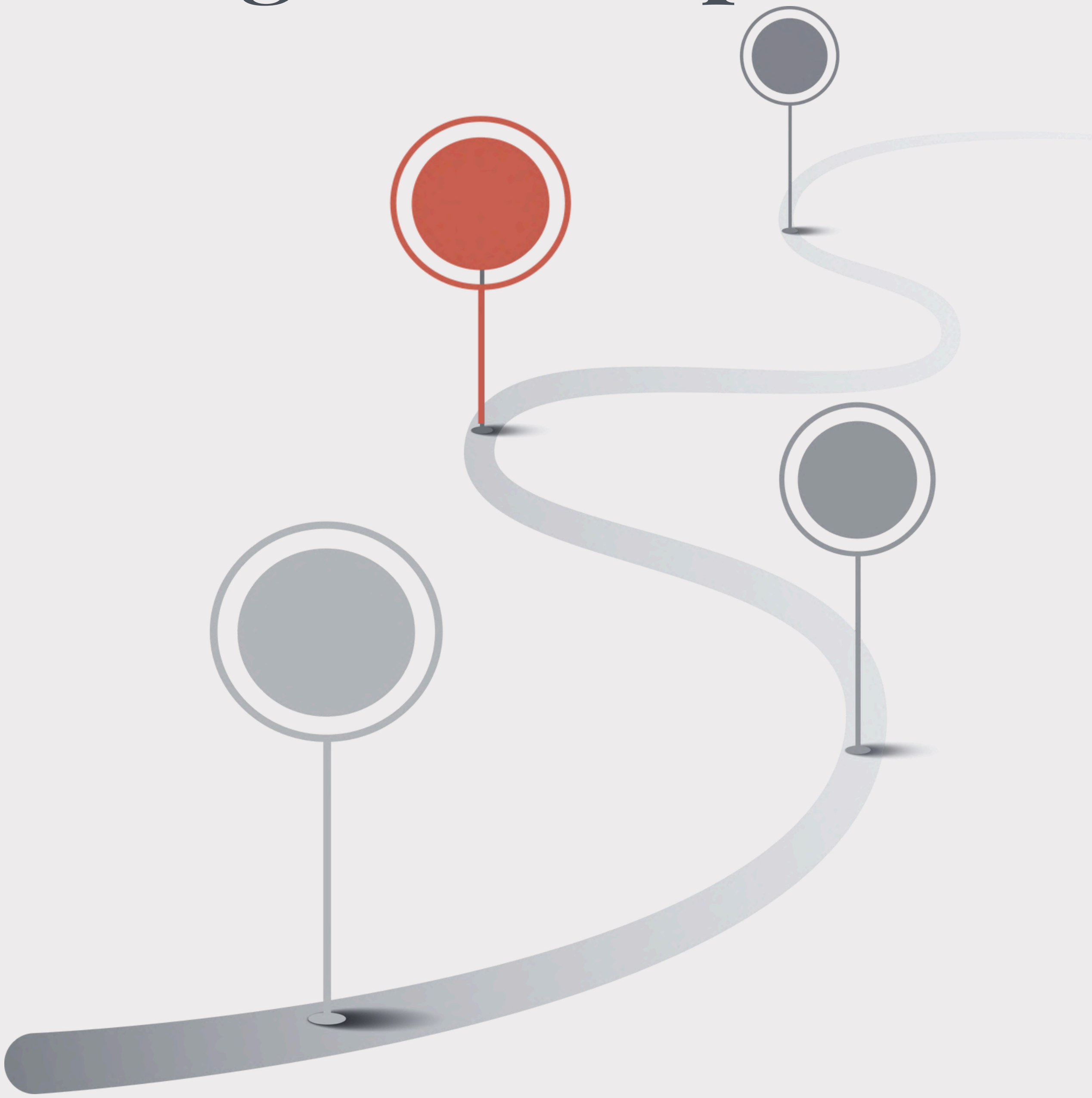
Image: Lakehouse Vision, The Oracle vs. The Architect

as needs evolve, building Lego-like stacks tailored to specific workloads.

Interoperability tools such as Apache XTable further enable seamless interaction across table formats without duplication or reconfiguration. The result is an adaptable architecture that supports evolving AI use cases, faster experimentation, and future-proof data platforms, moving enterprises away from rigid systems toward architectures that can evolve with business and AI demands.

Q2

Architecting New Design Principles

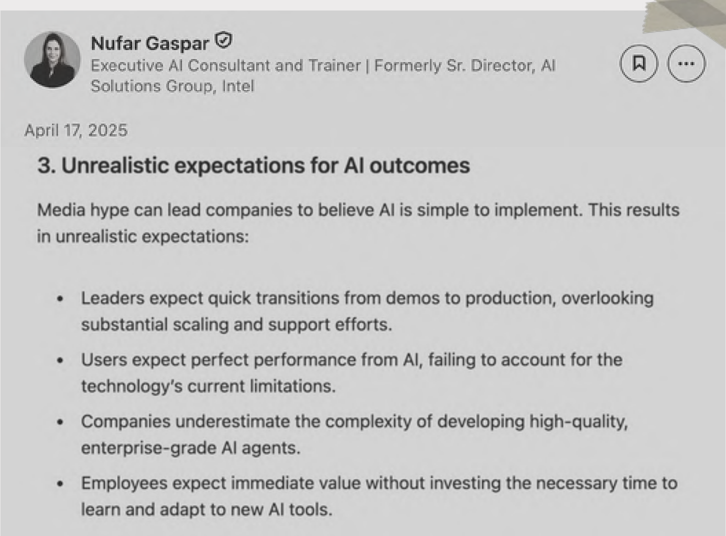


Q2: The Fragility in AI Adoption

Crumbling Data Stacks & PoC Purgatory

2025 triggered a wake-up call across the industry where rapid AI adoption collided with brittle data foundations and operational overload. A report of 230+ practitioners across 48 countries, with an average of 15+ years of experience exposed 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. (source)

“With AI dominating headlines, executives feel the pressure and rush to implement AI solutions because they feel left behind — not because they've identified specific problems AI can actually solve,” mentions Sol Rashidi, Chief Strategy Officer (CSO), AI & Data at Cyera. (Source)


Nufar Gaspar
Executive AI Consultant and Trainer | Formerly Sr. Director, AI Solutions Group, Intel

April 17, 2025

3. Unrealistic expectations for AI outcomes

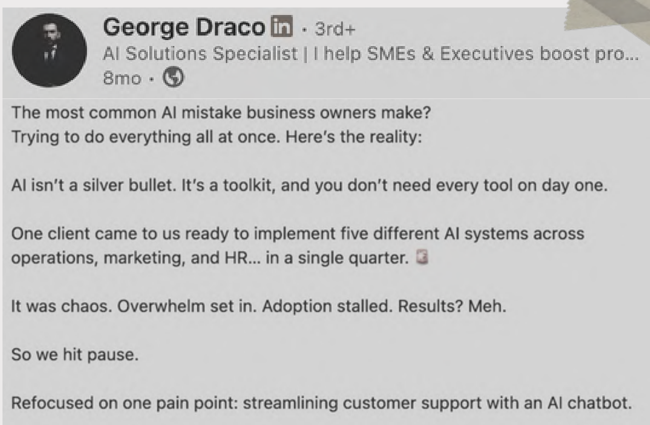
Media hype can lead companies to believe AI is simple to implement. This results in unrealistic expectations:

- Leaders expect quick transitions from demos to production, overlooking substantial scaling and support efforts.
- Users expect perfect performance from AI, failing to account for the technology's current limitations.
- Companies underestimate the complexity of developing high-quality, enterprise-grade AI agents.
- Employees expect immediate value without investing the necessary time to learn and adapt to new AI tools.

What actually bothered real AI success was its velocity without real returns.

Tool sprawl, integration friction, siloed AI pilots, and PoC purgatory traced back to a single root cause: data and governance were excluded from the acceleration loop.

The core failure wasn’t experimentation, but the inability to move PoCs into production due to missing ownership, quality standards, success metrics, and guardrails. Adoption increased, but organisational maturity, measurement, ROI tracking, and platform readiness lagged behind.


George Draco · 3rd+
AI Solutions Specialist | I help SMEs & Executives boost pro...
8mo ·

The most common AI mistake business owners make?
Trying to do everything all at once. Here's the reality:

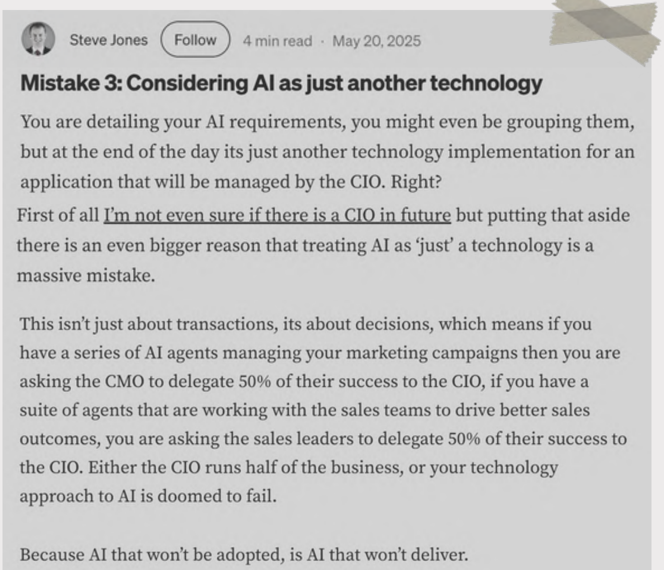
AI isn't a silver bullet. It's a toolkit, and you don't need every tool on day one.

One client came to us ready to implement five different AI systems across operations, marketing, and HR... in a single quarter.

It was chaos. Overwhelm set in. Adoption stalled. Results? Meh.

So we hit pause.

Refocused on one pain point: streamlining customer support with an AI chatbot.


Steve Jones Follow 4 min read · May 20, 2025

Mistake 3: Considering AI as just another technology

You are detailing your AI requirements, you might even be grouping them, but at the end of the day its just another technology implementation for an application that will be managed by the CIO. Right?

First of all I'm not even sure if there is a CIO in future but putting that aside there is an even bigger reason that treating AI as ‘just’ a technology is a massive mistake.

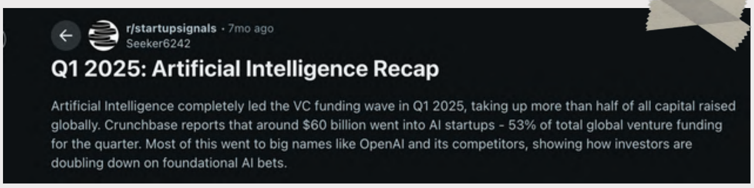
This isn't just about transactions, its about decisions, which means if you have a series of AI agents managing your marketing campaigns then you are asking the CMO to delegate 50% of their success to the CIO, if you have a suite of agents that are working with the sales teams to drive better sales outcomes, you are asking the sales leaders to delegate 50% of their success to the CIO. Either the CIO runs half of the business, or your technology approach to AI is doomed to fail.

Because AI that won't be adopted, is AI that won't deliver.

Shift in Investor Sentiments

Investor behaviours have had their own trends of 'ups & downs' in 2025, with a good boost initially and drops or unpredictable moves in the later parts.

In the first two quarters, AI startups commanded over half of global venture funding, but reports highlighted a bifurcation where established AI players attracted disproportionate capital while early-stage or non-AI startups struggled for attention.



r/startupsignals · 7mo ago
Seeker6242

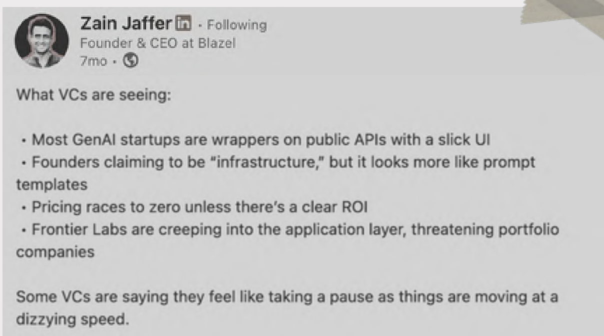
Q1 2025: Artificial Intelligence Recap

Artificial Intelligence completely led the VC funding wave in Q1 2025, taking up more than half of all capital raised globally. Crunchbase reports that around $60 billion went into AI startups - 53% of total global venture funding for the quarter. Most of this went to big names like OpenAI and its competitors, showing how investors are doubling down on foundational AI bets.

What was the trend signalling? A new VC magnet could be AI-native, enterprise-grade SaaS products. A June 2025 VC survey noted that while AI disruption was seen as rising, investor optimism was cooling, with a growing share of VC respondents scrutinising deals more closely and pausing or tightening investment criteria amid broader uncertainty and competitive pressures (source).

Investors Saw Early Signs of "AI FOMO" Driving Funding

Analysis of early funding patterns highlighted that capital was flowing disproportionately into AI startups, especially late-stage rounds, suggesting a fear of missing out rather than disciplined selection based on fundamentals.



Zain Jaffer · Following
Founder & CEO at Blazel
7mo ·

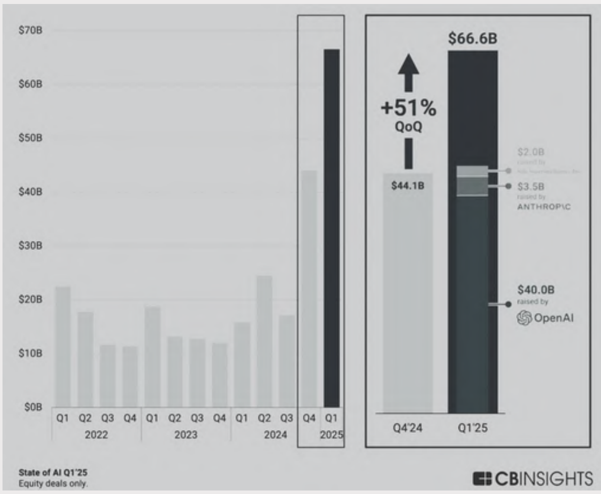
What VCs are seeing:

- Most GenAI startups are wrappers on public APIs with a slick UI
- Founders claiming to be "infrastructure," but it looks more like prompt templates
- Pricing races to zero unless there's a clear ROI
- Frontier Labs are creeping into the application layer, threatening portfolio companies

Some VCs are saying they feel like taking a pause as things are moving at a dizzying speed.

While PitchBook's survey data focuses on broader tech, it also highlights that investor expectations of AI impact increased sharply in key domains, and AI remained central in portfolios. At the same time, barriers to adoption, such as unclear use cases and high costs, had become top concerns, signalling that investor optimism was tempered by realism about where value would actually materialise.

Image: Outlook for the second half of the year: enthusiasm is high, but caution remains | Source



Metadata Returned to the Center of AI Trust

The GenAI wave exposed these gaps, shifting the conversation from model capability to rebuilding trust through better data. Metadata is re-emerging not as documentation, but as the scaffolding for governance, reuse, and explainability.
Early GenAI efforts often skipped lineage, policy enforcement, and data quality in favour of speed. As the costs of blind automation surfaced, metadata moved back to the centre as a design principle.

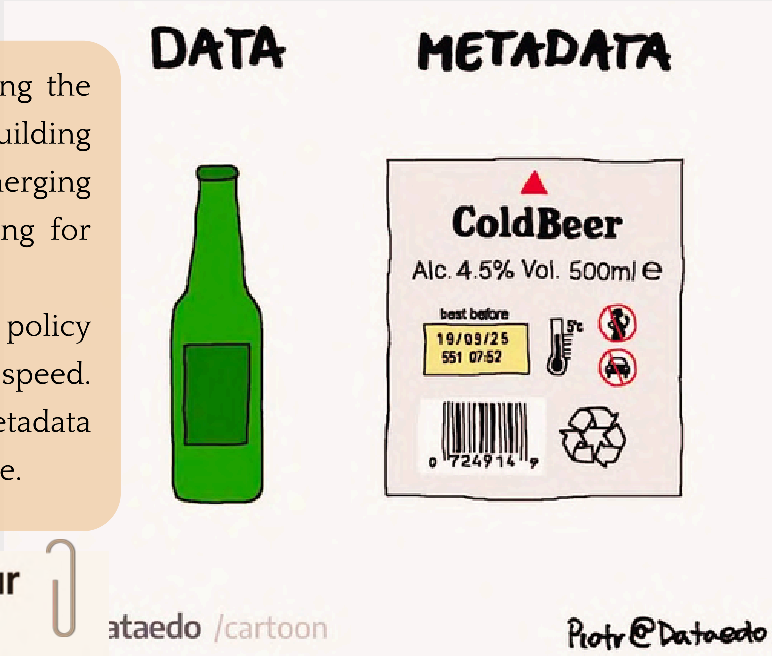


Image: The difference between data & metadata | Source

Metadata: The New Lifeline for Your GenAI Initiatives

Julia Bardmesser
Helping Companies Maximize the Business Value of Data and AI | ex-CDO advising CDOs at Data4Real | Keynote Speaker &...

All of a sudden, the very things that used to feel like "data plumbing" – context, consistency, source tracking – have become make-or-break factors in whether your AI works well or goes off the rails.
GenAI will confidently give you wrong answers if it doesn't understand the context of your data.

Think about it: when you ask an AI system about "customer churn," does it know you're referring to voluntary cancellations versus involuntary account closures?

Without semantic context - what we data geeks call business metadata - your GenAI models start hallucinating or returning results that are irrelevant, or worse, incorrect.

That semantic context lives in metadata.

Bradley Shimmin · 2nd
Global Technology Analyst | Focused on helping companies...
7mo ·

+ Follow

Informatica World 2025: It's always been about the metadata

Fast forward to this morning's Informatica World keynote, where CEO Amit Walia echoed this same view...at first. It did not take long, however, for Mr. Walia to take a sharp turn toward metadata. Yes, we're talking about data about data, with companies increasingly building a solid semantic layer of meaning, access, and control across their disparate and fractured data estates. In short and paraphrasing, he said that companies won't succeed in building agentic AI through data alone. Rather, they will succeed or fail depending on the quality of the data about that data.

What's Holding Back AI In Your Organization? One Word: Metadata.

Raman Tallamraju
Technology Executive | Data, AI & Cloud Strategy | Head of Enterprise Data Architecture, Vanguard

June 1, 2025

Metadata is how we give data its voice. It is not merely description, it is interpretation. A machine does not "see" data the way we do. Metadata is the lens through which both humans and machines come to understand data consistently.

The Rise of Memory-Centric AI Architectures

Inside the Shift from Context to Continuity

As AI systems moved from experimentation to execution, one limitation has become impossible to ignore: memory.

As AI systems moved from experimentation to execution, one limitation has become impossible to ignore: memory. LLMs can generate and reason, but they lack persistent, structured memory. This is where the idea of a Memory Operating System (MemOS) is gaining attention.

A system like MemOS manages memory for LLMs through persistent, structured memory units with versioning, governance, and lifecycle management. Rather than treating memory as an ephemeral context or cache, it frames memory as a first-class, governable asset. In this model, AI agents rely on memory that can be queried, updated, audited, and reused across tasks and sessions.

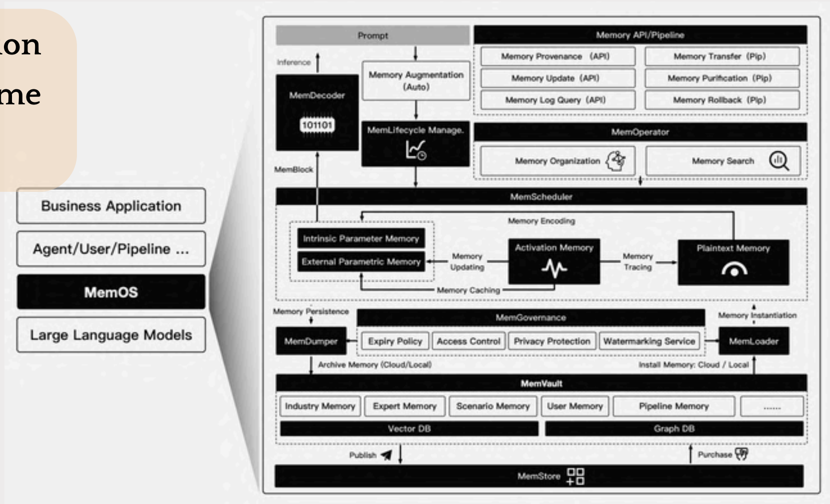


Image: Overview of the MemOS architecture (Source)

What distinguishes MemOS from traditional vector databases or cache-based approaches is its focus on agentic readiness. Memories are not passive logs; they are actionable and API-accessible. Components such as indexing, summarisation, retrieval, and consistency checking allow agents to learn continuously and operate coherently over time.

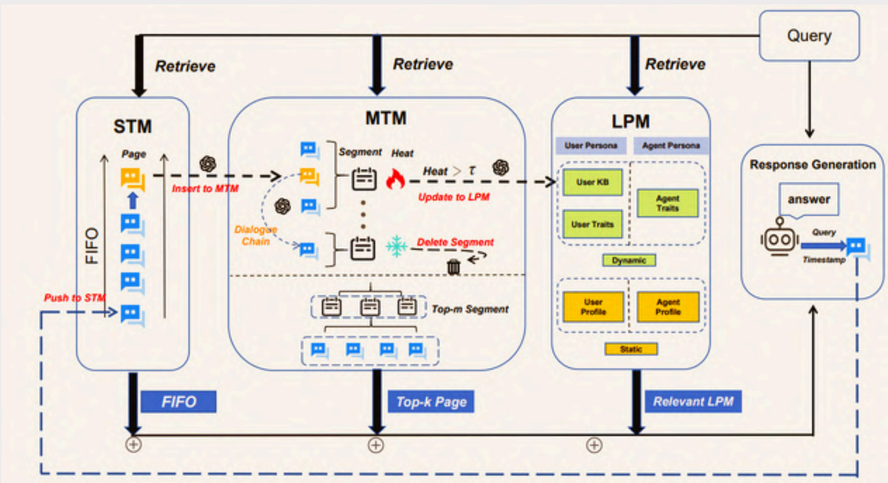


Image: The overview architecture of MemoryOS, including memory Store, Updating, Retrieval, Response | Source

This enables advanced use cases, autonomous customer support agents, legal copilots, and enterprise advisors that must remember prior interactions, decisions, and outcomes.

Anchoring the Agentic Stack for Continuity

After a year defined by acceleration, more models, more agents, more pilots, the industry is entering an anchoring phase.

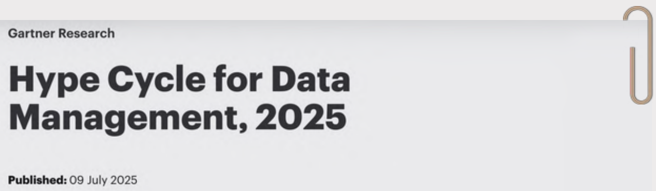
Progress is no longer measured by speed alone, but by whether AI systems can retain context, explain decisions, and evolve safely over time.

This is where operating-system thinking enters the stack. Concepts like Memory OS and Data OS signal a shift from isolated capabilities to foundational layers that orchestrate data, context, and memory as shared utilities.

“However, prevailing architectures remain anchored in static parameters and lack structured modeling and unified management of memory, rendering them inadequate for supporting knowledge updates, state retention, and personalized adaptation.”

~ an insight from paper on MemOS.

Rather than treating memory as an add-on, these platforms frame it as infrastructure: governed, reusable, and continuously learning across agents, workflows, and interfaces.


Gartner Research

Hype Cycle for Data Management, 2025

Published: 09 July 2025

“Embracing an Operating System for Data does not have to mean a radical or disruptive transformation of your data environment.

Implementing an Operating System for data such as DataOS refines your data philosophy, allowing you to harness and operationalize your data more effectively without the heavy lifting often associated with tooling changes or migration.”

-an excerpt from DataOS Philosophy.

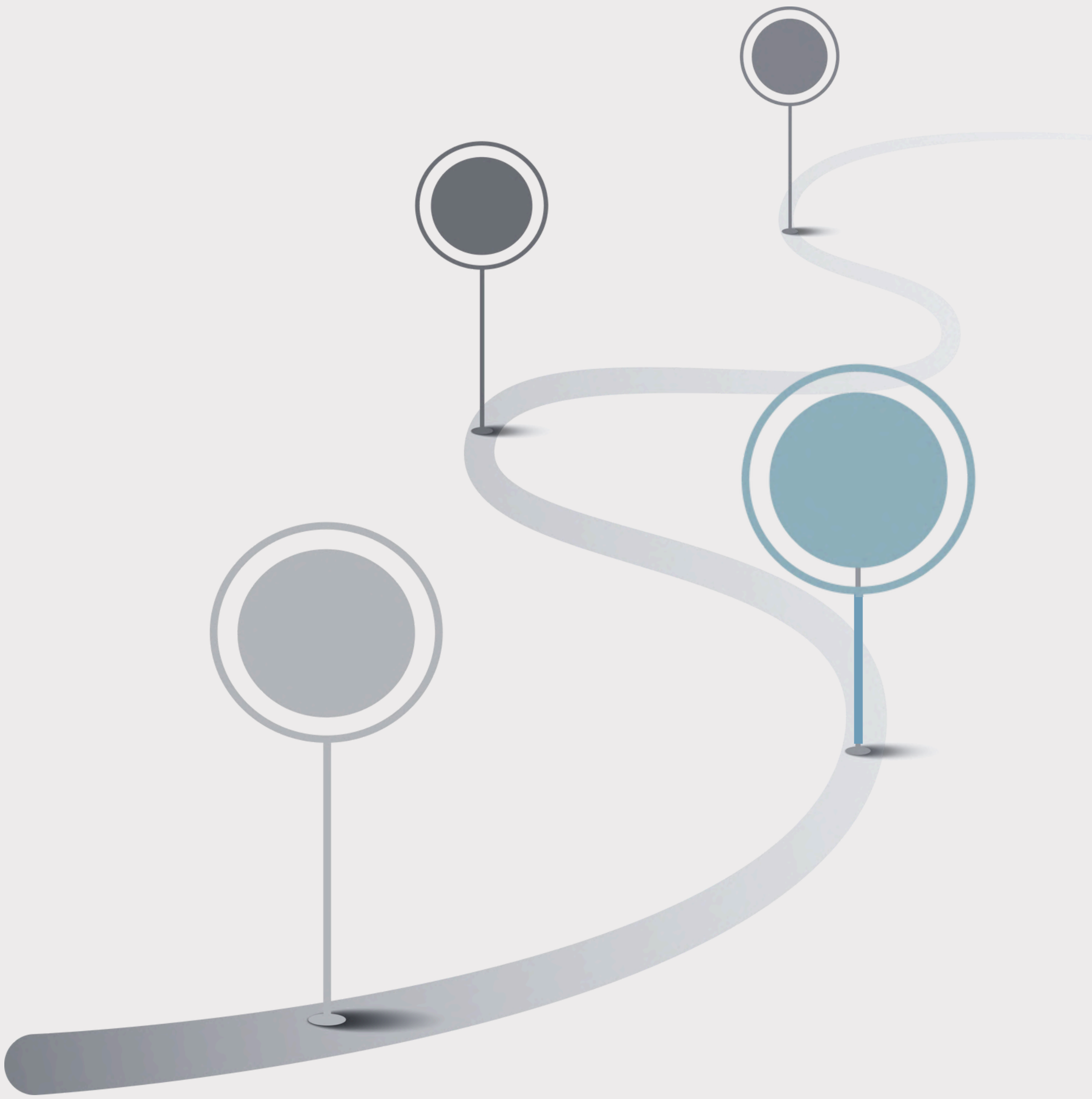
Importantly, this shift is no longer theoretical. Gartner’s Hype Cycle (by Aaron Rosenbaum, Senior Research Director at Gartner, & Robert Thanaraj, Senior Director at Gartner), has formally recognised this direction, identifying DataOS by The Modern Data Company among the solutions anchoring the next wave of data architecture.

Other recognised vendors include Agile Lab, IBM, Informatica, K2view, Nexla, One Data, Revelate, and Starburst, signalling broad industry alignment around this architectural evolution.

The signal was clear. The future of AI will not be defined by faster models alone, but by systems built on trusted data, persistent memory, and shared context. Intelligence may drive capability, but continuity is what makes it sustainable.

Q3

Re-preparing for AI



Data Platforms Hit Their Limits

Enterprise platforms reached a breaking point in Q3. Teams realised they had optimised for technical capability, not usability or adoption. Feature-rich platforms stalled because users were instructed to use them, not drawn to them. As data volumes, domains, and real-time demands multiplied, older platform designs, built for simpler contexts, collapsed under complexity.

Dylan Anderson · 1st
The Data Ecosystem Author ✦ Head of Data Strategy @ Profusion/ Atom...
View my blog
5mo ·

They say you want you can't have

With data, its more like leadership wants what they don't understand...

The technology and approach to setting up their data platform is front and centre, especially with terms like Mesh and Fabric that sound sexy, but are incredibly hard to understand

Christopher Pike · 2nd
Engineer
8mo ·

Your data platform "works"—so why isn't anyone using it?

One of the most persistent problems in data engineering is that we often fail to tie our work
to the actual goals of the business. The classic symptom is a dashboard that no one uses. My
data platform was a more expensive and less visible example of the same class of failure.
Defining a model that binds data platform implementation to business needs can help avoid that, while fostering better collaboration between technical and business partners.

Sagar Paul
SVP Global Growth & Solutions at The Modern Data Company

"The goal of a data platform is not to capture the present, but to anticipate the future and build a schema of possibility. Today's design should already know where tomorrow's needs will fit, even if those needs are not yet visible."

The Data Platform Is a Contradiction: Why AI Demands Something Better

Paul Gale
Driving Automation & Innovation in Data Lifecycle | Platform Engineering | Data Strategy & Governance Advocate

September 2, 2025

The Problem With the "Data Platform"

The typical Data Platform is often built separately from a developer workflow and product delivery and run by a central team responsible for ingestion, pipeline, governance, and access control.

Most platforms remained reactive: answering requests, building dashboards, responding to queries. They lacked intent-driven feedback loops where systems could evolve alongside business logic. Context stayed siloed, governance broke at scale, pipelines grew brittle, and AI initiatives starved for meaningful, connected data.

As Peter Baumann noted in one of his articles, many architectures still mirror the legacy systems they replaced: cloud-hosted silos stitched together after the fact. The flaw wasn't distribution or tooling; it was design. Until semantics and relationships are treated as first-class citizens, platforms will continue managing data instead of activating it.

Paolo Platter · 1st
CTO - Co-founder at Agile Lab | Product Manager on Witboost
Visit my website
4mo ·

📊 Most data platforms scale infrastructure. Very few scale decision-making.

As the number of data teams grows, the challenge isn't just provisioning environments or automating pipelines.

It's ensuring that:
- Everyone knows what good looks like
- Practices evolve without breaking trust
- Change is codified and visible, not just tribal
- Governance scales without centralizing everything

This is the real platform maturity curve:
➡️ From automating tasks
➡️ to shaping behaviors
➡️ to enable change at scale.

Shifting Focus from Data Access to Context Engineering

tobi lutke
@tobi

I really like the term "context engineering" over prompt engineering.

It describes the core skill better: the art of providing all the context for the task to be plausibly solvable by the LLM.

Avivah Litan · 2nd
VP and Distinguished Analyst at Gartner Inc.
5mo · Edited ·

‼ Context engineering is essential to success in agentic AI systems. It addresses and lowers high failure rates — often around 60% to 90% in complex multiagent systems (MAS) that arise from interagent misalignment and coordination breakdowns, according to industry studies.

Oren Grinker · 3rd+
Co-Founder & VP R&D @ Jeen.AI
3mo ·

While prompt engineering dominated conversations in 2023, a more fundamental challenge has emerged: context engineering—the discipline of providing AI models with exactly the right information, in the right format, at the right time.

ADDY OSMANI
JUL 14, 2025

engineering could run. In other words, a witty one-off prompt might have wowed us in demos, but building **reliable, industrial-strength LLM systems** demanded something more comprehensive.

This realization is why our field is coalescing around "**context engineering**" as a better descriptor for the craft of getting great results from AI. Context engineering means constructing the entire **context window** an LLM sees – not just a short instruction, but all the relevant background info, examples, and guidance needed for the task.

Aaron Levie · 3rd+
CEO at Box - Intelligent Content Management
5mo ·

Context engineering is increasingly the most critical component for building effective AI Agents in the enterprise right now. This will ultimately be the long pole in the tent for AI Agents adoption in most organizations.

Shifting Focus to Semantics & Context Engineering

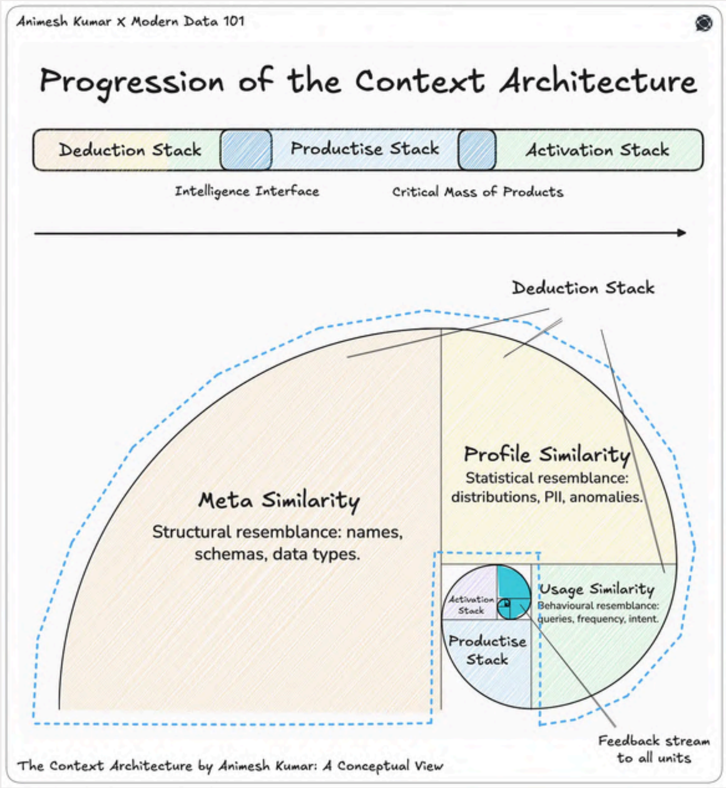


Image: Progression of the Context Architecture | Source: Animesh Kumar

As AI systems shift from accessing data to understanding it, context becomes the real source of power. Most AI pilots stall as intelligence without context cannot scale. Context architecture closes this gap by turning disconnected data into explainable, usable, and decision-ready intelligence.

"The absence of context boundaries means that every table, dashboard, and metric floats in isolation. They are fragments of truth without a shared narrative." mentions Animesh Kumar, in his article, Rise of the Context Architecture

Traditional pipelines strip away business context during transformation, leaving machines to learn from technically correct but semantically void data, causing AI outputs that are statistically sound yet misaligned with business intent. A semantic layer embedded within data products preserves a consistent business context across system changes. It translates raw codes into business concepts and ensures continuity when underlying systems evolve, dramatically reducing the friction and fragmentation that typically bog down AI implementations.

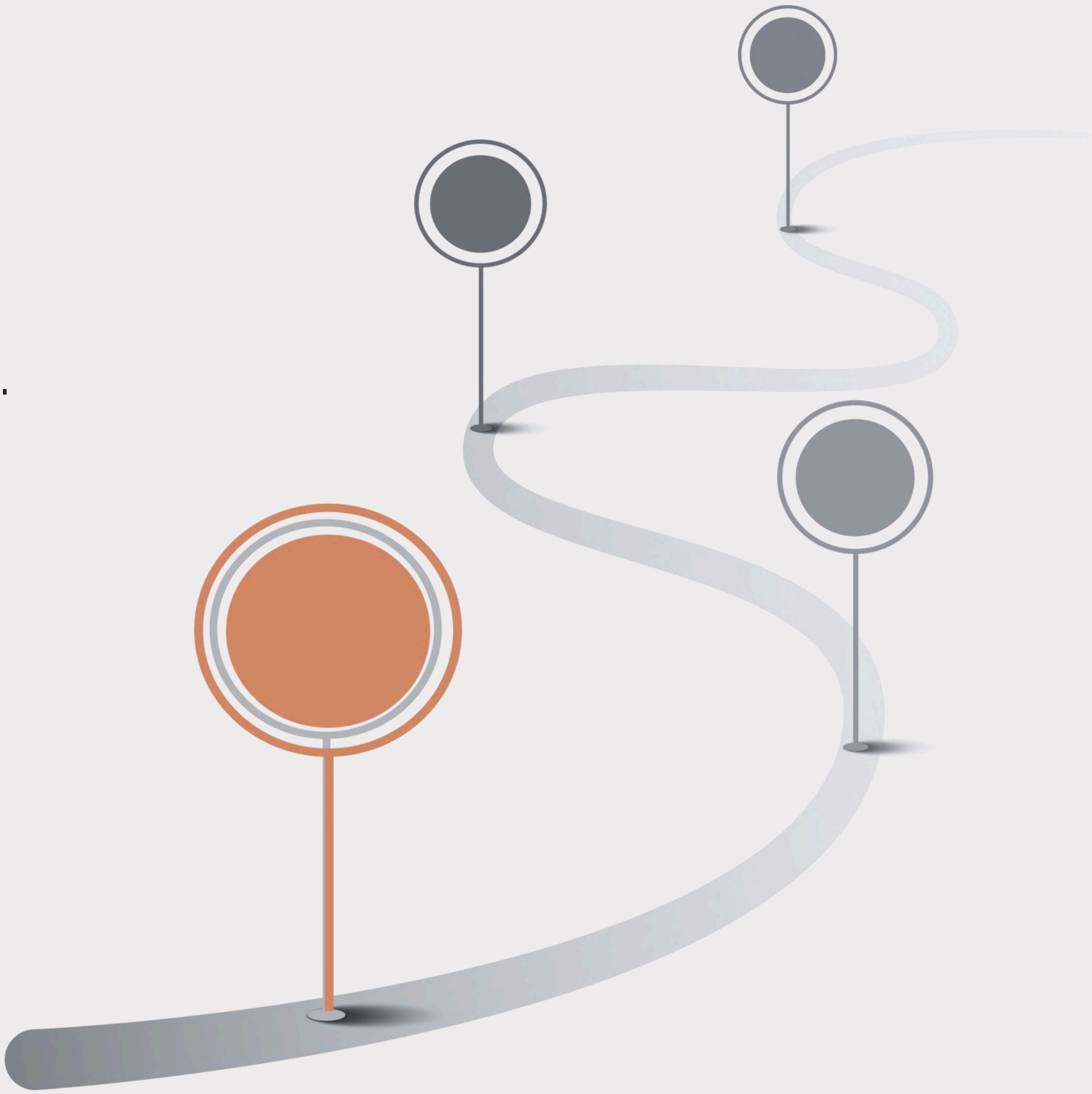
The emerging consensus is that traditional semantic models, built primarily for BI, are insufficient for GenAI and agentic systems. What's taking shape instead is a next-generation semantic layer that unifies data, information, and knowledge management into a shared operational plane, one where meaning and context can be consistently understood by humans, analytics tools, and AI agents alike.

"A next gen semantic layer unifies the worlds of data management, information management, and knowledge management," mentions Malcolm Hawker, CDO at Profisee (Source)

"It will also allow for companies to apply structure to all of the insights buried in their highly unstructured data sources, and to govern this data at scale."

Q4

Data Becomes the Control Plane



Governance Debt Became an Operational Risk

All that hard work towards building the right platform and semantic layers often led to unexpected failures, and definitely, the engineering wasn't at fault. They were some ignorance, some shortcuts!

Though basic governance was ignored until it broke trust. Governance deferred ("we'll document later," "hardcode it for now") eroded trust until it surfaced as conflicting metrics, silent AI failures, and audit dead-ends.

Marcel Dybalski · 1st
AI & Data Platform Engineer | GCP & Fabric Certified
2mo ·

"Just add one more column."

Sure. Let me also:
→ Rebuild the entire data pipeline
→ Recalculate all historical metrics
→ Test 47 edge cases you didn't mention
→ Update documentation nobody reads
→ Break 3 downstream reports in the process

One column = 40 hours of work.

This is why data teams are drowning.

Every "small ask" assumes the infrastructure is already perfect.
But most organizations are running on duct tape and prayer.

No data dictionary.
No governance.
No single source of truth.

Just vibes and Excel formulas from 2016.

John Wernfeldt · Following
CDOs & data leaders → from firefighting to strategic authority | Governa...
Visit my website
3mo ·

Most data platforms don't collapse because of technology.
They collapse because of what's missing: governance.

Here's the pattern I keep seeing:

→ New data source? "Just hook it up."
→ New metric request? "Hardcode it for now."
→ Lineage questions? "We'll document later."
→ Ownership issue? "Ask around."

On the surface, everything looks fine. Dashboards load. Queries run. Leadership claps.
But underneath, the wall is crooked. Every new brick makes it worse.

Nicola Askham
DataIQ 100 2022 | Award Winning Data Governance Training | Consultant | Coaching | Data Governance Expert | D.A.T.A Foundin...

Let me start by addressing the elephant in the room: traditional data governance is reactive. We've built entire frameworks around fixing problems *after* they occur, rather than preventing them from happening in the first place.

Here's what this typically looks like in practice:

- Data gets created and pushed to production without proper metadata
- Data stewards scramble to document and classify assets after the fact
- Quality issues are discovered downstream when reports fail or decisions go wrong
- Data governance teams spend their time in endless catch-up mode, always one step behind

This approach creates what I call the "governance gap"—the dangerous space

Clare Kitching · 2nd
Transform your AI & data ambition into action | xQ
1mo ·

Data governance gets labeled as red tape.
Here's why it's actually your organisation's unsung he

What people think it is:
→ Writing rules
→ Limiting access to data
→ Endless red tape
→ Saying "no"

Carolyn Birchill · 2nd · 1mo
Modernizing Data for AI | Governance, Quality and Operations Leaders...

For me the most overlooked part of data governance is what is going on quietly in the background. Because when it's working well, any use of data whether it's reports or AI runs smoothly. At the moment I'm leaning into the storytelling of that and bringing these good news stories to light so the value can be understood and therefore ensuring funding continues.

How Leaders Shifted Focus to Data Lineage Gaps

Lineage gaps, unclear ownership, and unmanaged changes turned governance debt into real operational risk. Q4 of 2025 observed how governance stopped being framed as compliance overhead and was finally recognised for what it is: the structural integrity that keeps AI systems stable, explainable, and safe to scale.

As organisations rely more heavily on analytics and AI, the data feeding these systems must be explainable, traceable, and trustworthy. Lineage ensures inputs are not only technically correct but also contextually understood, supporting operational confidence and meeting rising regulatory expectations around AI transparency and accountability.

A recent conversation led by Ian Barker, a tech journalist & occasional novelist, highlighted the crucial question of: **Why is data lineage particularly important for AI?**

“The greatest value that a data lineage creates is a shared understanding across teams,” mentions Tom Baeyens, Co-founder and CTO at Soda.

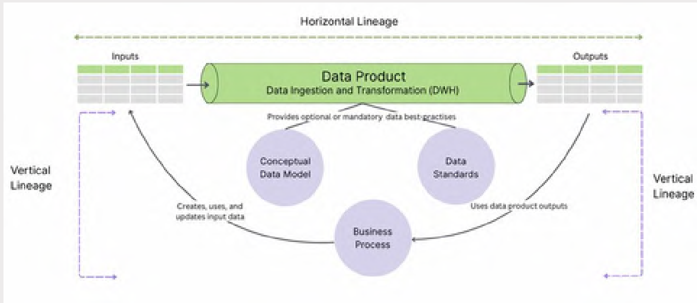
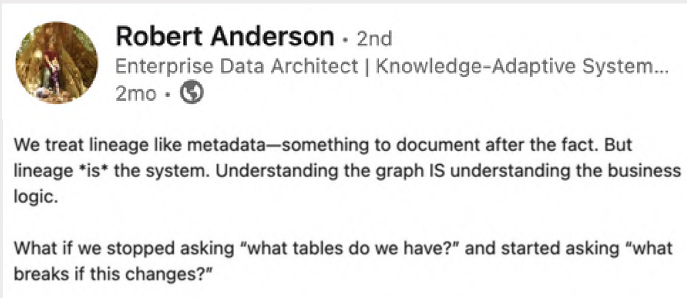


Image: How horizontal and vertical lineage look across the data product | Source



Effective lineage requires shifting from passive reconstruction to proactive design. By organising data around well-owned, context-rich data products, lineage becomes embedded across the data lifecycle. Treated strategically, lineage enables trust, speed, and scale, turning AI, governance, and reuse from bottlenecks into competitive advantage rather than technical debt.

> Unlike traditional analytics, where a human reviews a dashboard before making a decision, AI systems operate autonomously. When your churn prediction model suddenly starts flagging loyal customers as high-risk, you need to quickly trace whether the issue stems from a pipeline change, a data quality problem, or a shift in business logic upstream ~ Saurabh Gupta, Chief Strategy & Revenue Officer at a Data Platform Company

How Leaders Shifted Focus to Data Lineage Gaps

Dr. Irina Steenbeek, Managing Director at Data Crossroads, in one her articles, points out aptly not just the importance of lineage today, but how regulation remains the strongest driver for lineage adoption. This article also highlights how organisations must maintain end-to-end, attribute-level lineage that stays accurate as systems, ownership, and business logic change

Data Lineage: Challenges and Trends 2025. Part 1: Why Data Lineage Matters —And How the Landscape Is Shifting

Dr. Irina Steenbeek
Strategic Advisor and Solution-Driven Coach for Data & AI Governance Frameworks | Maturity Assessment | Data Lineage |...

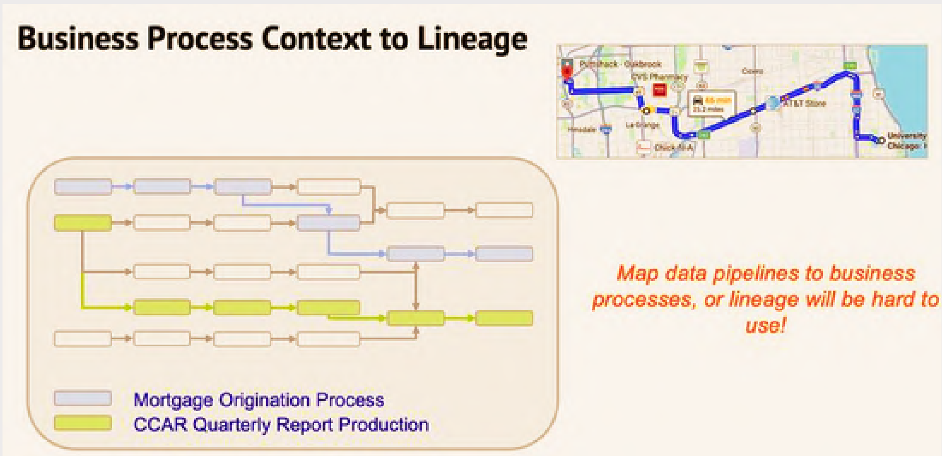
Business Drivers Behind Lineage Adoption

Regulatory Compliance and Risk Transparency

The most powerful driver is regulation. Supervisory bodies now expect organizations—especially in the financial sector—to provide full transparency into their risk data. The **BCBS 239 principles**, expanded through the **ECB 2024 supervisory guide**, outline detailed expectations for what data lineage must include.

linking technical flows to business terms and governance structures. Lineage must be embedded in formal data governance, connected to data quality validation, managed within a governed metadata environment that supports monitoring, accountability, and version control. The image shows Data Lineage Documentation, Five-Year Trends (Source).

The real value of data lineage lies in shared understanding. In a post by Olga Maydanchik, Director Data Analytics at Anywhere Real Estate Inc., she presents a



unique approach to lineage and states how data lineage only becomes usable when it is anchored in a business context. Using a Google Maps analogy, the author shows that capturing every road (tables, pipelines, transformations) isn't enough. Users need a guided path that highlights what matters for a specific business process.

Rising Concern of Strategic Jargons

Circling Back to the Gap Between Expectation vs. Realtity for AI

The industry continues to offer plenty of sophisticated terms for data & AI developments that ultimately become goals for enterprises. However, a lot of them are just marketing jargon and buzzwords that have no specific means of assessment.

Think of AI readiness this way: it is contextual, dependent on use case, environment, and business needs, and not a universal state. Many organisations still think data is ready for AI, but readiness assessments reveal serious gaps in quality, governance, and access, showing the difference between belief and reality. Unfortunately, surveys and analysis show that poor data quality continues to be a top risk for AI success, yet companies treat it as a checkbox rather than a nuanced, use-case-specific requirement. Much prior to the year ending, Andreas Horn, Head of AIOps | EMEA at IBM, pointed out.

"There's a lot of noise around AI right now: buzzwords like "agentic AI", "AGI" and "Fine-tuning" are everywhere! It's easier than ever for people to present themselves as experts without having any technical foundation or practical experience to back it up."

True expertise bridges technical depth with business strategy, while recognising AI's wider societal impact. If someone can't explain the fundamentals clearly, they're likely amplifying hype rather than driving meaningful insight or innovation.

Malcolm Hawker · 1st
CDO | Author | Keynote Speaker | Podcast Host
3mo ·

What does 'AI Ready Data' really mean?

For three years, a litany of vendors, consultants, and thought leaders have been telling us repeatedly that bad data leads to bad AI.

Yet somehow, ChatGPT was built on top of a data quality cesspool.

The phrase 'bad data = bad AI' is what's known as a platitude.

Platitudes are factually unassailable, but also largely meaningless. (See also, 'garbage in, garbage out')

You can't act on a platitude because it lacks specificity, and as such, it cannot withstand any real interrogation.

Yet in our world of data and analytics, we (ironically) rely heavily on them.

Shaun McGirr · 2nd 3mo
I bring practical experience to bear on your AI transformation

I have started calling such platitudes "lullabies"

Like · 2 | Reply · 4 replies

Industry veterans did have similar opinions on this, especially when it comes to concepts like "AI-ready" or "well-governed" without being able to explain for which use case, at what level of risk, or against which decision.

Platitudes reduce nuanced realities to simplistic binaries, and this undermines credibility and usefulness.

When Data Stopped Being Just an Infrastructure

Leadership's Move of Turning Data Into Business Strategy

One of the most strategic moves emerged when institutional alignment started catching up with practitioner reality.

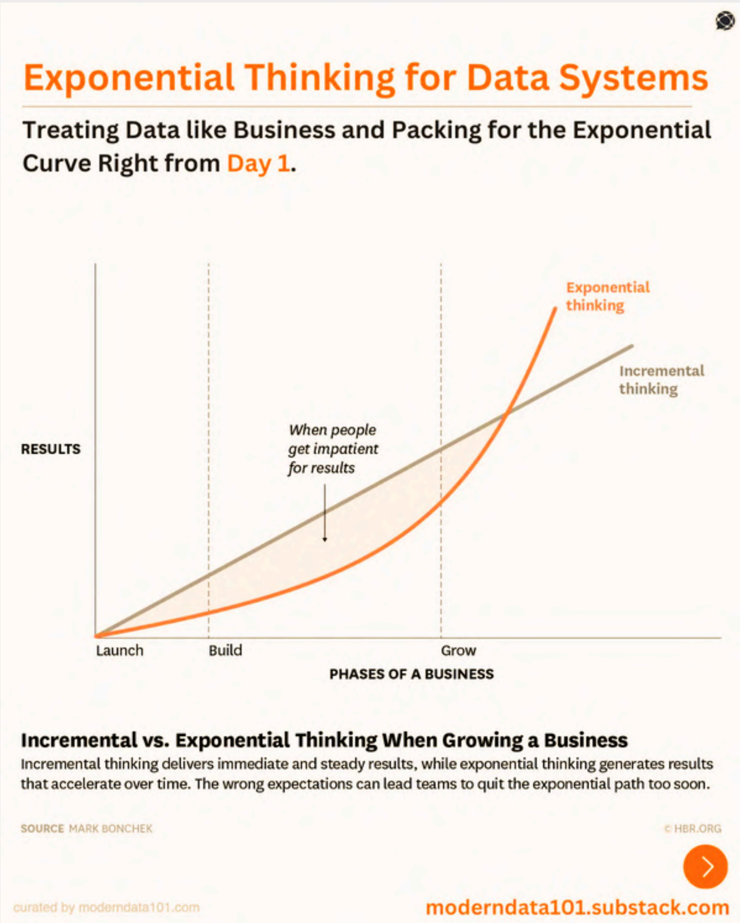


Image: Exponential Approach in Data Systems | Source

By Q4, enterprise leadership formally caught up with what practitioners had been signalling all year: AI and data are no longer technical enablers, they are strategic levers.

A shift was reinforced by growing acknowledgement that agentic AI demands architectural change, not incremental tooling. At the same time, the CDO's role visibly evolved from custodian to strategist, with data positioned as a driver of revenue, decision velocity, and competitive advantage.

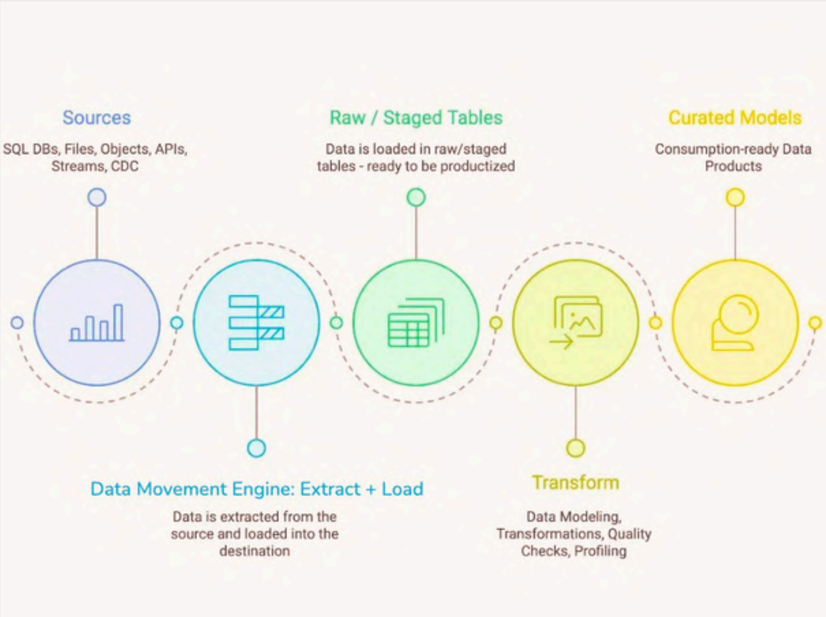
"The mindset shift required: data as a strategic asset. To scale AI, leaders must stop seeing data as an application byproduct. Work backwards: define business outcomes first, then identify the critical workflows and data needed," mentions Michał Krzyżanowski, Head of Data and AI at KUBO.

As data moved into the strategy conversation, leaders increasingly confronted a clarity gap: business outcomes were expected, but few organisations had reliable measures to link data and AI investments to revenue, efficiency, or risk reduction. Hence, strategic intent without measurable value quickly dissolves executive confidence.

.

When Data Stopped Being Just an Infrastructure

Designing Data Movement Around Time, Cost, and Trust



Data Movement Engine Fit | Source

Data movement has been treated as plumbing, an invisible, technical concern solved by pipelines, schedules, and retries. In 2025, that assumption finally broke.

As organisations push data products into operational and AI-driven use cases, they are discovering a hard truth: most data products fail because the way data moves no longer matches how it is consumed.

The core insight is simple but disruptive: Data movement is about preserving state, trust, and timeliness under real-world constraints.

The First-Principles Reset

Modern data movement must be designed around three non-negotiable constraints:

Time: How fresh does the data need to be for decisions, products, and AI agents?

Cost: What compute and operational overhead is acceptable at scale?

Trust: Can consumers rely on the data being complete, ordered, and correct?

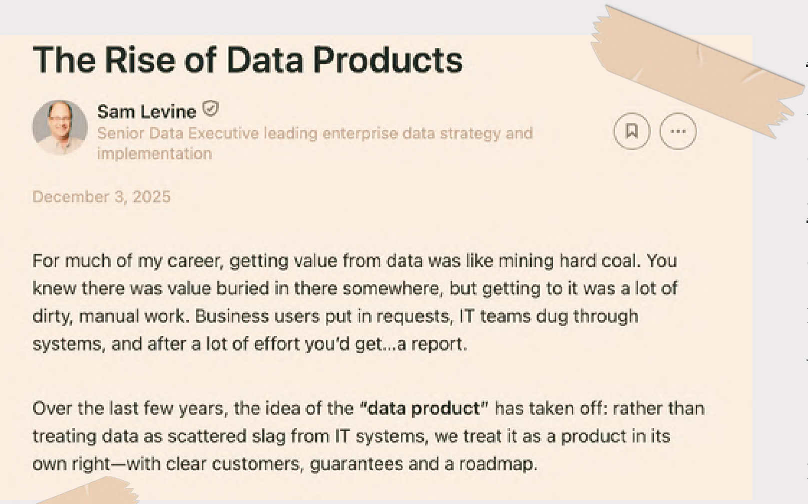
What's Changing in Practice

Teams that are successfully operationalising data products are converging on a new movement model.

Instead of full reloads, only what changes moves forward; deltas become the default, with CDC treated as a baseline requirement for correctness and low latency.

APIs are no longer edge cases but first-class, state-bearing sources alongside databases. Trust is designed in, not retrofitted, through built-in observability that makes movement inspectable and explainable in real time. And critically, extensibility moves to the edges: teams closest to the data can onboard sources and destinations independently, without central platform bottlenecks.

Relooking at Data Products for Improved Business Alignment



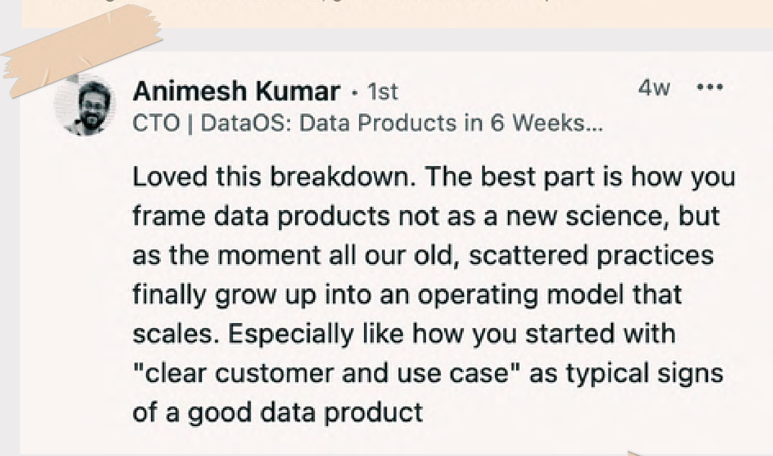
The Rise of Data Products

Sam Levine
Senior Data Executive leading enterprise data strategy and implementation

December 3, 2025

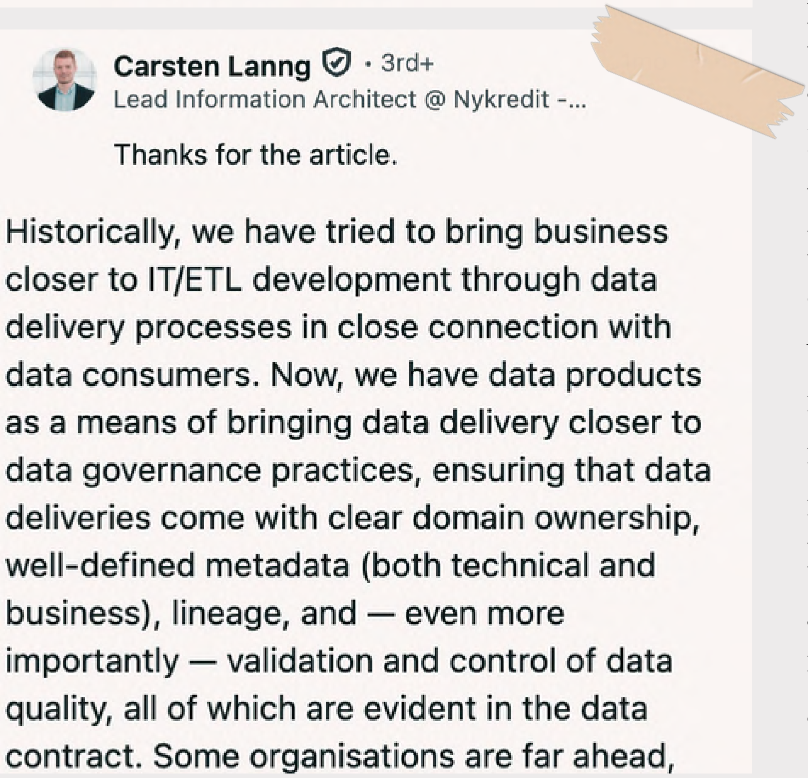
For much of my career, getting value from data was like mining hard coal. You knew there was value buried in there somewhere, but getting to it was a lot of dirty, manual work. Business users put in requests, IT teams dug through systems, and after a lot of effort you'd get...a report.

Over the last few years, the idea of the **"data product"** has taken off: rather than treating data as scattered slag from IT systems, we treat it as a product in its own right—with clear customers, guarantees and a roadmap.



Animesh Kumar · 1st · 4w
CTO | DataOS: Data Products in 6 Weeks...

Loved this breakdown. The best part is how you frame data products not as a new science, but as the moment all our old, scattered practices finally grow up into an operating model that scales. Especially like how you started with "clear customer and use case" as typical signs of a good data product



Carsten Lanng · 3rd+
Lead Information Architect @ Nykredit -...

Thanks for the article.

Historically, we have tried to bring business closer to IT/ETL development through data delivery processes in close connection with data consumers. Now, we have data products as a means of bringing data delivery closer to data governance practices, ensuring that data deliveries come with clear domain ownership, well-defined metadata (both technical and business), lineage, and — even more importantly — validation and control of data quality, all of which are evident in the data contract. Some organisations are far ahead,

Mahdi Karabiben, Senior Product Manager at Neo4j, offered some crucial insights in one of his masterclasses into how organisations can get this right. He focused on the fundamentals of data modelling, which is essential in the data product space because it gives data shape, meaning, and governance in ways that mirror real business value.

"Modelling starts when you start listening," says Mahdi Karabiben, Senior Product Manager at Neo4j.

He believes that effective data modelling is foundational to creating scalable, business-aligned data products, and it must start with business context, meaningful metrics, and governance from day one.

Rather than treating modelling as a technical afterthought, models should be anchored in real business use cases, structured by domain, and built with clear ownership and evolving metrics. Ultimately, data products require models that reflect business reality, and features like semantic layers, metric trees, and iterative governance are essential to ensuring models remain adaptable, transparent, and useful across teams and use cases.

Why Marketplaces Are Taking Center Stage

Marketplaces have surely not gone out of importance, but at this crucial point of the AI momentum, marketplaces have become the key to enhanced data adoption. When products are ready to drop into analytics or AI workflows with embedded governance and semantic context, time-to-insight accelerates significantly and trust compounds with every reuse. Priyanshi Durbha, Principal, Advanced Analytics, aptly highlighted this in a session on democratising data for humans and AI through data marketplaces.

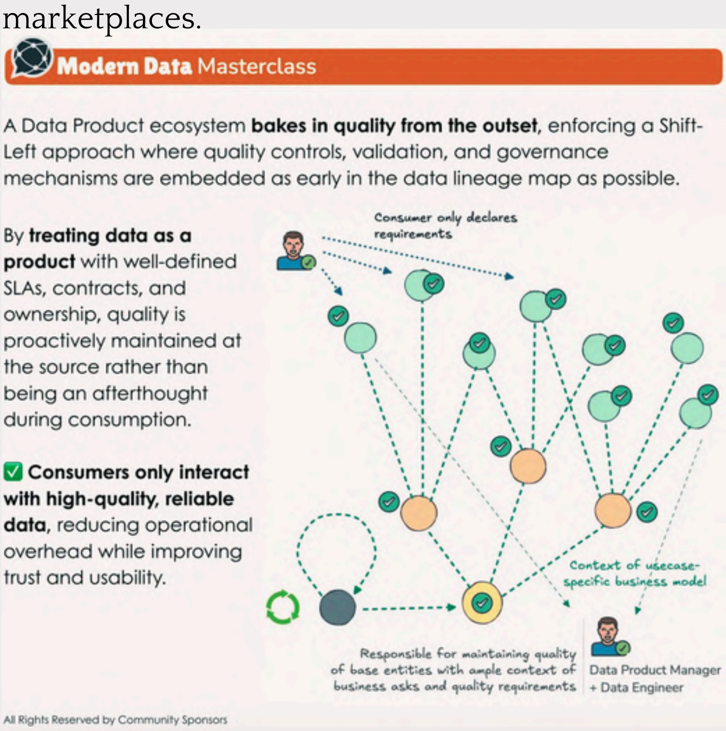


Image: How Data Product Marketplace alters Consumption Ease and Patterns | Source: Masterclass: Boosting Data Adoption with Data Product Marketplaces

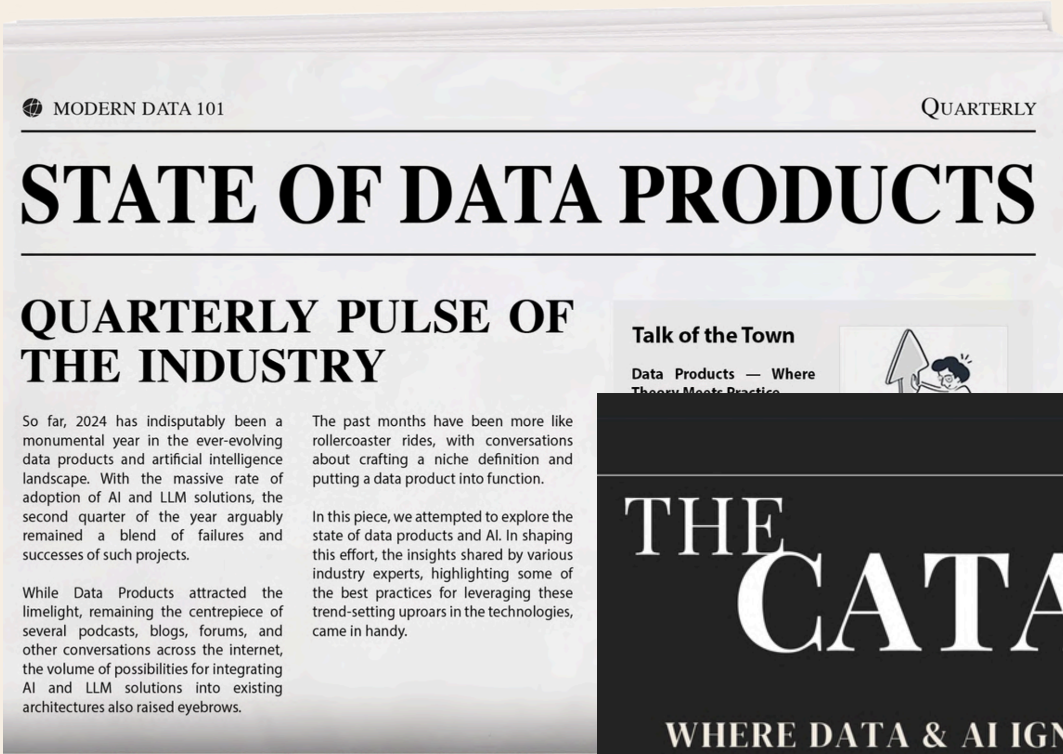
"Data marketplaces are quickly becoming essential for any organization trying to scale AI. They turn scattered data assets into governed, trusted, ready-to-use data products: all accessible through a single, intuitive experience," mentions Randy Rouse, Field CTO at Quest Software.

By turning discovery into intuitive exploration and embedding explainability into data products themselves, marketplaces increase adoption not by demanding more skills, but by making data intrinsically usable.

"Importantly, marketplaces serve both humans and AI agents. They provide user interfaces for human browsing and APIs for programmatic discovery and access. This dual capability is essential in the AI era," mentions Noemi Moreno Fabelo. Head of Data Strategy & Architecture at Cambridge University Press & Assessment.

Traditional data projects stall because teams spend excessive time finding, preparing, and validating data. A marketplace reframes adoption as data accessibility, making relevant data products easy to find and use.

Data and AI for Business
with State of Data Products


MODERN DATA 101 QUARTERLY

STATE OF DATA PRODUCTS

QUARTERLY PULSE OF THE INDUSTRY

So far, 2024 has indisputably been a monumental year in the ever-evolving data products and artificial intelligence landscape. With the massive rate of adoption of AI and LLM solutions, the second quarter of the year arguably remained a blend of failures and successes of such projects.

While Data Products attracted the limelight, remaining the centrepiece of several podcasts, blogs, forums, and other conversations across the internet, the volume of possibilities for integrating AI and LLM solutions into existing architectures also raised eyebrows.

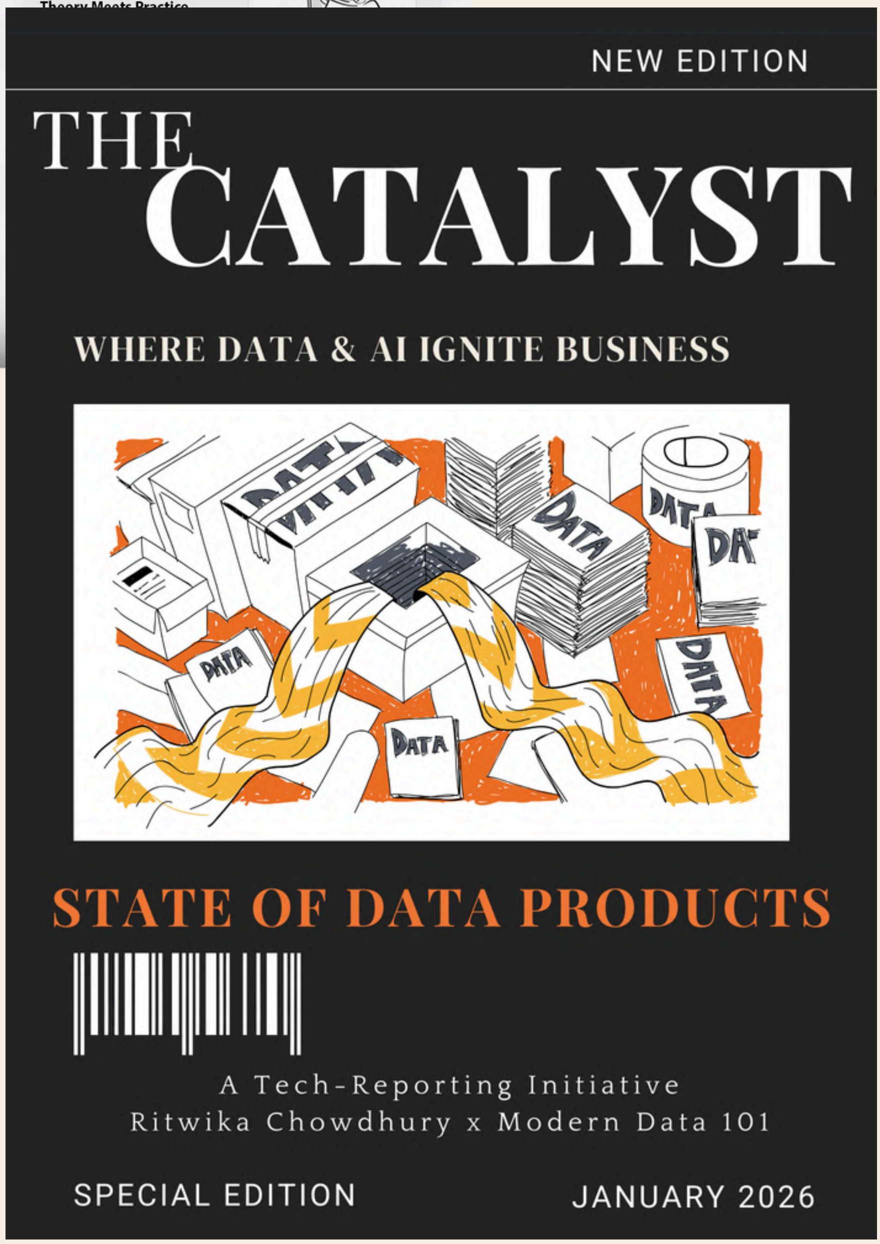
The past months have been more like rollercoaster rides, with conversations about crafting a niche definition and putting a data product into function.

In this piece, we attempted to explore the state of data products and AI. In shaping this effort, the insights shared by various industry experts, highlighting some of the best practices for leveraging these trend-setting uproars in the technologies, came in handy.

Talk of the Town

Data Products — Where

WE ARE IN OUR NEW COVER!



Modern Data 101

Join the Collective

moderndata101.com

Modern Data 101

THE CATALYST

THE COMMUNITY PULSE FOR DATA & AI

STATE OF DATA PRODUCTS

SPECIAL EDITION

JANUARY 2026