Modern Data 101 Q2 2026 EDITION

THE CATALYST

Semantic or Sorry?

STATE OF DATA PRODUCTS

A Tech-Reporting Initiative

Ritwika Chowdhury x Modern Data 101

QUARTERLY RELEASE: Q2 APRIL-JUNE

Founding Author's Note

THE CATALYST

SEMANTIC OR SORRY

This issue is packed with the innovations and newly emerged concepts that disrupted the data & AI landscape in Q2 of 2026.

Almost every enterprise I speak to is running AI at some layer or another. However, most carry a doubt about whether it is running well. There are so many gaps between activity and correctness, between adoption and production, between systems that are connected and systems that agree on what something means. This issue maps these gaps and talks about the next state.

The answer, more often than not, is cross-functional meaning and how it changes shape across functions, tools, and consumers. **Ontology** is the infrastructure that tells your AI what your business means; what a customer is, what revenue counts, what approved means in your specific context. Without it, every agent you deploy is guessing. And at scale, that guess does not stay small but compounds into liability.

What has struck me most in the conversations shaping this issue is how this realisation is arriving through a slow accumulation of agents that ran correctly by every operational measure and reasoned incorrectly by every business one.

Usually, observability stacks will tell you the agent ran, but not whether it reasoned well. That is the gap no monitoring tool or dashboard is closing right now, and it will define which enterprises can genuinely trust their AI to act, and which ones are simply monitoring failure at speed. Shared meaning, governed agents. Honest observability. That is the standard. This issue holds the industry to it.

Animesh Kumar

Q2 / APR–JUN 2026
THE CATALYST

THE CATALYST

SEMANTIC OR SORRY?

Modern Data 101
State of Data Products
Quarterly Edition

002

SEMANTIC CLARITY

The Real Bottleneck in Enterprise AI

It's not compute. It's meaning.

ONTOLOGY MATTERS

From Ambiguity to Accuracy

Explicit meaning. Reliable AI outcomes.

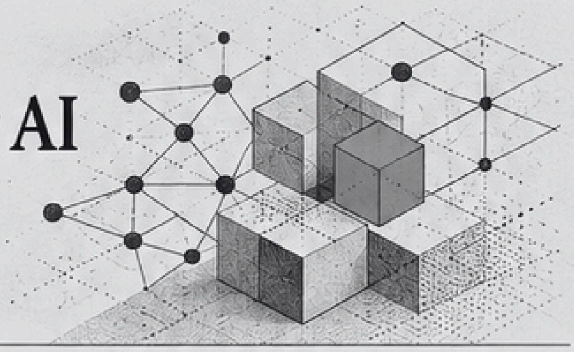
AI READY INFRASTRUCTURE

The Ontology Layer in the AI Stack

Bridging structured data and generative intelligence.

Why Ontology Is the Foundation of Trustworthy AI

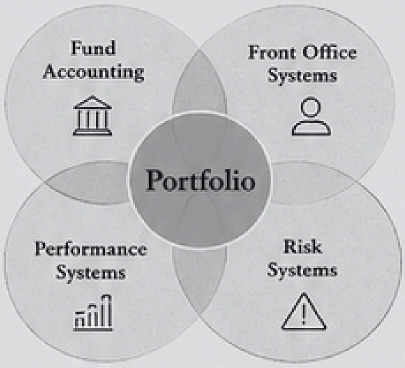
In enterprise AI, the real bottleneck is not model capability. It's semantic clarity. Ontology turns business meaning into machine-readable context, helping AI reason with accuracy, not just probability.



The Semantic Fragility of Investment Data

One word. Many meanings. Each system is right — in its own context.

- **Fund Accounting** – Legal fund vehicle with NAV & compliance
- **Front Office** – Strategy sleeve or mandate
- **Performance** – GIPS reporting composites
- **Risk Systems** – Exposure groupings for stress testing



Same word. Different meaning. AI gets confused.

The Risk Without Ontology

LLMs retrieve based on similarity, not meaning.

- ✕ Semantic blending across different concepts
- ✕ Numerically plausible but wrong answers
- ✕ Hallucinations? Not always. Often it's semantic confusion.

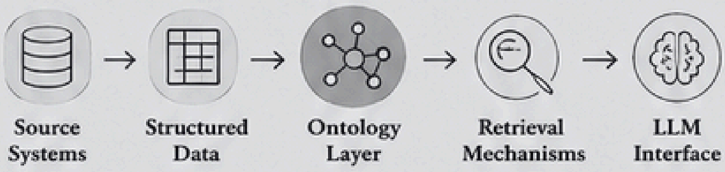
Result:
Confident answers built on ambiguous foundations.

What Ontology Actually Delivers

Ontology formally models business concepts, relationships, constraints, and context — so AI can reason within defined boundaries, not guess beyond them.

- **Clarity** – Defines entities, relationships, and hierarchies
- **Consistency** – One meaning, used everywhere
- **Accuracy** – Narrower reasoning space for AI
- **Governance** – Contextual rules and constraints enforced
- **Interoperability** – Systems speak the same semantic language
- **Trust** – AI outputs grounded in business reality

The Architectural Shift



Ontology is the bridge between enterprise data and generative reasoning.

Why It Matters More in 2026

- ✓ Agentic AI acts across systems without human ambiguity resolution.
- ✓ Fluent AI can mask semantic errors, making the cost of ambiguity higher.
- ✓ Enterprises need meaning, not just better models.
- ✓ Governance must extend from data quality to meaning quality.
- ✓ The differentiator is not who has the best model, but who has the clearest semantic foundation.

> *"Before asking if we are ready for generative AI at scale, a more fundamental question is: Have we defined the entities we operate with?"*

IN BRIEF

From Documentation
Ontology is evolving from documents to operational infrastructure.

From Guesswork
AI reasoning moves from similarity to defined business meaning.

From Activity To Reliability
Moving from AI outputs that look right to ones that are right.

From Hype To Foundation
Shared meaning is the foundation of trustworthy enterprise AI.

The Ontology Gap Is Becoming the AI Gap

Let’s address the first question first: Why were enterprises functioning without ontologies in the first place, or rather, ‘how?’

Ontologies weren't skipped out of negligence; they were unnecessary by design.

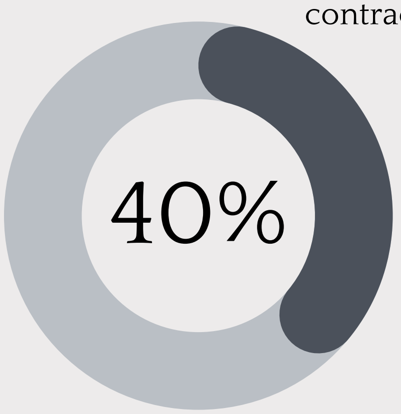
Ontology has gone from a niche knowledge-management concern to non-negotiable infrastructure for agentic AI. The argument: LLMs and agents now act without continuous human oversight and ambiguity that humans used to silently resolve (is "revenue" GAAP-recognised or bookings? is "customer" the CRM record or the billing record?) becomes a hard failure point for autonomous systems.

As Stephen Kaufman, Chief Architect, Americas CTO Office at Microsoft mentions in his article,

“This worked because meaning was localized and lived in silos. Each system encoded its own definition of a “customer,” “policy,” or “contract,” often embedded in application logic, SQL, business rules or in people’s heads.”

Research in knowledge graphs, semantic technologies, and explainable AI has consistently shown that explicit domain knowledge improves reasoning, interoperability, and traceability.

The latest wave of agentic AI extends this principle, positioning enterprise ontologies as operational infrastructure that grounds AI decisions in business context rather than probabilistic inference alone. As autonomous agents begin making decisions across business functions, ontologies are emerging as the shared semantic contract that enables consistent reasoning across data, applications, and workflows.



of enterprise applications will be integrated with task-specific AI agents by the end of 2026 (Source: Gartner);

each of those agents needs a shared semantic structure, or it picks whichever definition it encounters first.. Enterprise architecture is no longer just about systems, applications and processes. In the age of autonomous AI agents, it is increasingly about maintaining a living semantic model that every agent can trust. The enterprise ontology is evolving from documentation into operational infrastructure.

The Ontology Gap In 2026: Why Engterprises Worry?

As AI shifts from generating content to taking actions, enterprises need explicit business meaning. Ontologies close this gap by grounding AI in domain knowledge, relationships, and governance, enabling trustworthy, explainable, and reliable decisions that embeddings and prompts alone cannot provide.

Image: How ontologies add semantic context, explicit relationships, and explainable reasoning to overcome the limitations of vector embeddings for trustworthy AI.

One of the pieces from Soumapdip De, AI Product Manager at The Modern Data Company, infact, largely focuses on the real need for ontologies, even over vector embeddings.

Vector embeddings retrieve similar information, but they don't understand business meaning. Ontologies provide explicit relationships, business context, governance, and reasoning, enabling AI to move beyond similarity search to trustworthy, explainable, enterprise-grade decision-making.

As AI shifts from generating content to taking actions, enterprises need explicit business meaning. Ontologies close this gap by grounding AI in domain knowledge, relationships, and governance, enabling trustworthy, explainable, and reliable decisions that embeddings and prompts alone cannot provide.

As Jessica Talisman wrote in one of her posts, how the AI context gap is fundamentally a semantic gap. Formal ontologies provide the business meaning, provenance, relationships, and reasoning that systems of record and LLMs lack, transforming fragmented data into trustworthy, explainable, and reusable enterprise knowledge.

> **What is missing is not tooling but the recognition that this is a knowledge management problem, requiring knowledge engineers, information architects, ontologists and the cultural conditions that allow their work to be valued and sustained as part of the operational infrastructure.**

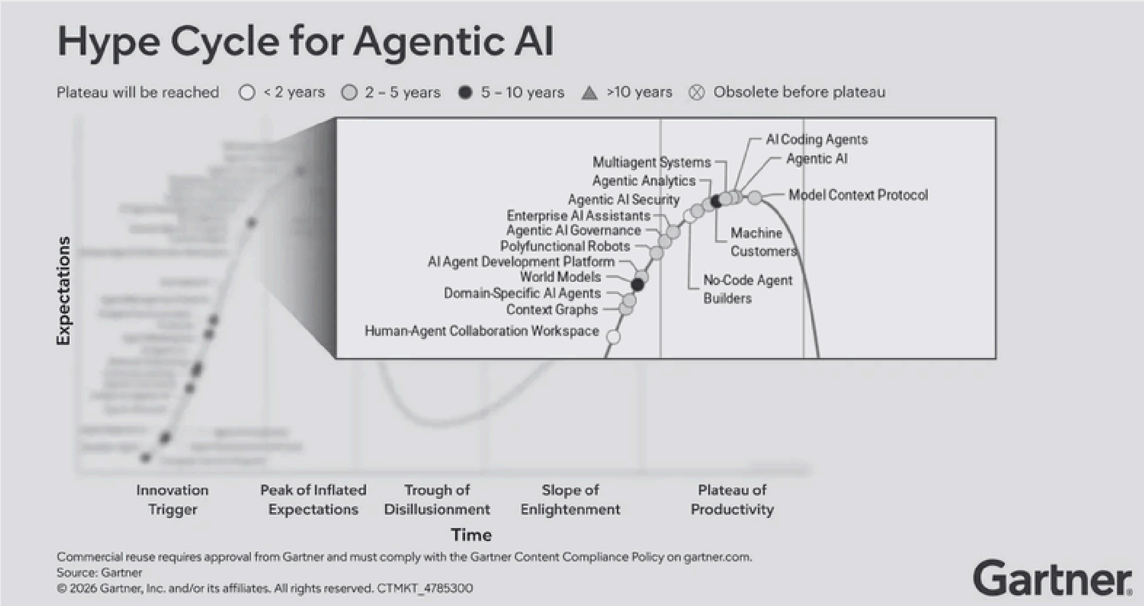
Quote by Jessica Talisman, Contextually LLC | Ontology Pipeline™| Knowledge Graphs

Enterprise Architecture Enters Its Semantic Era

Vijosh Ammayath in • 3rd+
Enterprise Field Engineering Leader | Cloud-Native Architec...
3mo • Edited •

What scales is simple:

1. Ontology before agents
Define canonical business objects first, then let every agent inherit the same business meaning.



> "Context with semantic coherence will become a cost-control and trust strategy, not a nice-to-have." – Rital Sallam, Distinguished VP Analyst at Gartner

Through ontologies, AI agents understand the framework of a particular problem, system or area of the business environment they operate in. With provenance, the AI agent will understand how to make decisions within that framework in alignment with the organization.

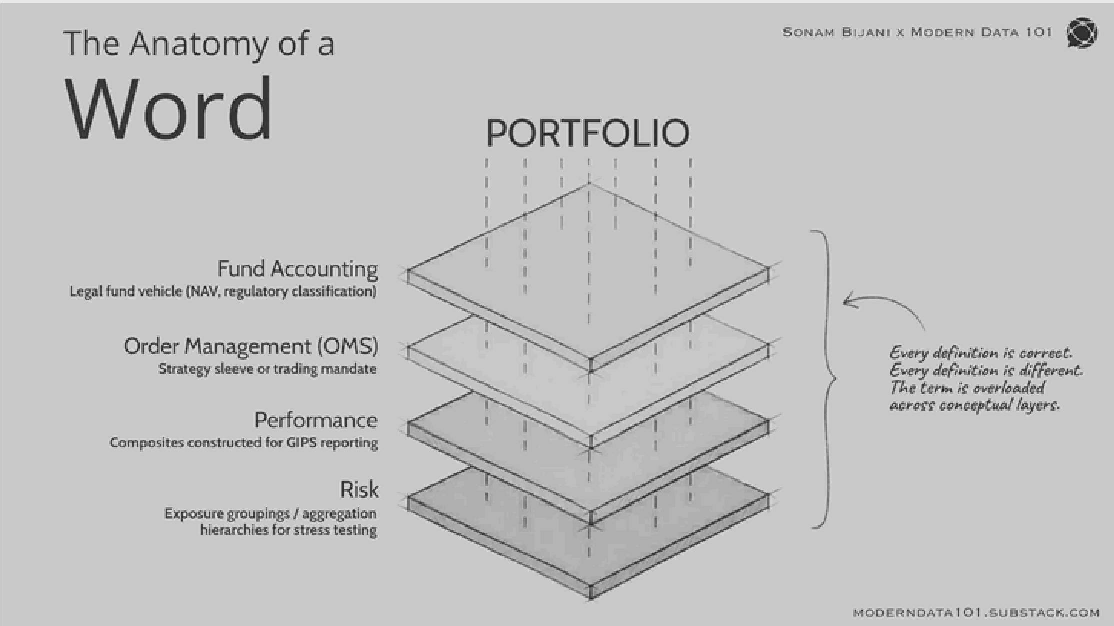
Ontology-Firsts or AI-Firsts?

AI-first made sense when the goal was proving the technology. You needed demos, you needed buy-in, you needed to show a board that the investment had a future. Ontology-first felt slow, theoretical, and hard to budget.

But AI-first has a ceiling. And most enterprises are hitting it right now. The agents fail not because the models are bad but because the models do not know what your business means. What is a customer? What counts as revenue? What is an approved supplier? Humans have been resolving that ambiguity silently for decades. Agents cannot. They guess, and they act on the guess.

Competitive timelines, board mandates, vendor urgency. Ontology was a later problem. What 2026 is revealing is that later has arrived, and it arrived faster than expected.

Years of pipeline-stacking without caring about meaning produced agents that fail precisely where semantic understanding was supposed to live.



It also describes the reality most enterprises are operating inside right now, AI is already running on an overhaul-required foundation, and the question has shifted from should we have built the foundation first to how do we stabilise what we have while continuing to ship. The sequencing debate does not resolve cleanly. What it does is sharpen the question every senior data leader should be asking right now: not "are we using AI" but "what does our AI think your business means, and is that what you actually mean?"

> The ontology narrows the hypothesis space within which the model can operate.

Quote by Sonam Bijlani, Data Quality Lead, Data & AI at **Eastspring Investments**.

Ontology gives AI explicit, machine-readable business meaning, allowing LLMs to reason over defined enterprise concepts instead of relying solely on statistical similarity, resulting in more accurate, trustworthy, and governable AI systems. Image: The Anatomy of a Word | Source.

Leaders say their data strategies will need a complete overhaul before their AI ambitions can succeed.

Enterprise AI & Data Products: "Data Grows Up"

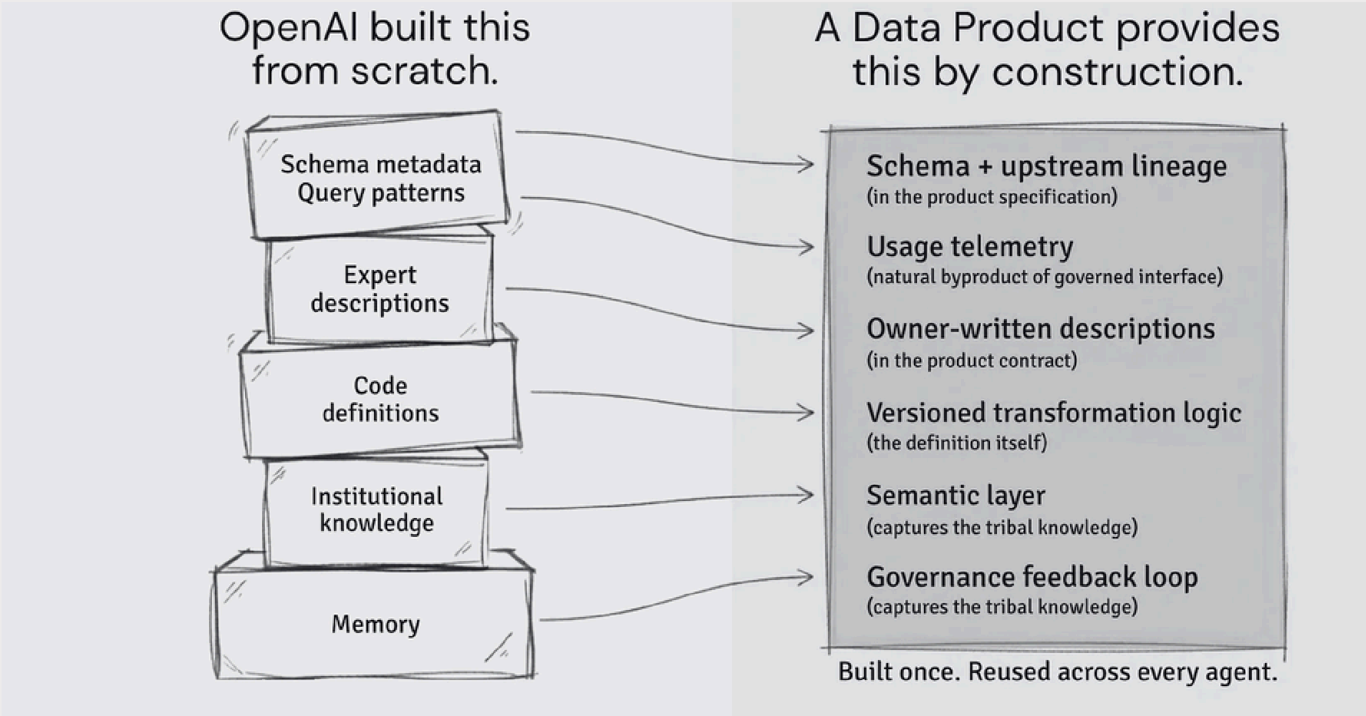
- Almost 70% of respondents say their data isn't clean or trustworthy enough for AI; 65% say it lacks the business context AI needs to be useful.
- Only 30% use AI for automated data preparation, despite data quality being the universally acknowledged bottleneck.
- 80% of respondents rank a semantic layer with standardised definitions as the most important enabler of AI — ahead of AI tools themselves and ahead of faster processing.

This is why data products stopped being a whiteboard idea. Animesh Kumar, CTO & Co-Founder, Shubhanshu Jain, Director of Product Management and Devendra Singh Rathore, VP of Engineering at The Modern Data Company, in one of their articles, argue "a Data Product is what an enterprise builds once, and agents consume forever", replacing agents rebuilding context from scratch per query (source).

Thoughtworks makes the same case from industry practice, calling owned, contracted data products "the essential foundation for success with trustworthy, production-grade AI" (source).

The honest failure mode: many "data domains" are rebadged org charts, still siloed under short-term IT budgets. Alejandro Aboy, Senior Data & AI Engineer at Workpath, warns that "a data product that fails silently is worse than one that fails loudly" (source), silent failure never generates a fix.

Image: What OpenAI built from scratch, in months, a Data Product provides by construction. The correspondence is near-exact - Source



Your Agent Has a Confession to Make

"People talk about agents as if the hard problem is building them. It isn't. The hard problem is verifying them."
mentions, Daniel Hulme,

He references 28,000 AI agents deployed across WPP Open and makes the case that most infrastructure discussions are missing the measurement and visibility layer entirely. Strong practitioner voice, not a vendor- exactly the expert type your audience respects.

Image: Agent Washing meaning | Source

Is Everyone ~~Brain~~ Agent-Washed

The AI Agent Spectrum: A Definitive Guide to What's Real, What's Rebranded, and What Actually Matters

Glen Cathey
Applied AI | Future of Work | Sourcing & Recruiting Expert | LinkedIn Learning & Social Talent Author

February 23, 2026

Every vendor in your inbox right now claims to have "AI agents." Your ATS provider has them. Your CRM has them. That startup that just raised $20 million definitely has them. There's just one problem - according to Gartner, only about 130 of the thousands of companies claiming to sell agentic AI are actually selling anything that qualifies.

Welcome to the era of ***agent washing*** - and it's going to cost companies that can't tell the difference.

David Linthicum · 2nd
Top 10 Global Cloud & AI Influencer | Full Stack AI Architect...
Visit my website
6mo ·

Agent Washing Is the New Cloud Washing — But This Time, the Stakes Are Higher

We've been here before.

The Truth About Enterprise Agentic AI in 2026: You Are Not Behind.

Paul Brzozowski
Enterprise AI ROI | Co-Founder @ Olakai.ai (Andrew Ng's AI Fund) | Proving AI Value Across Agentic, Assistive & Coding AI | AI System...

Deloitte's most recent Emerging Technology Trends study reports that only 11 percent of organizations are actively running agentic AI systems in production. Another 14 percent have solutions ready to deploy. Meanwhile, 42 percent are still developing their agentic strategy roadmap, and 35 percent have no formal strategy at all.

Read that again. Eighty-nine percent of organizations are not running agents in production.

Sri Narasimhan
Head of Strategy | AI & Enterprise Growth Strategist | Advisor & Board Member

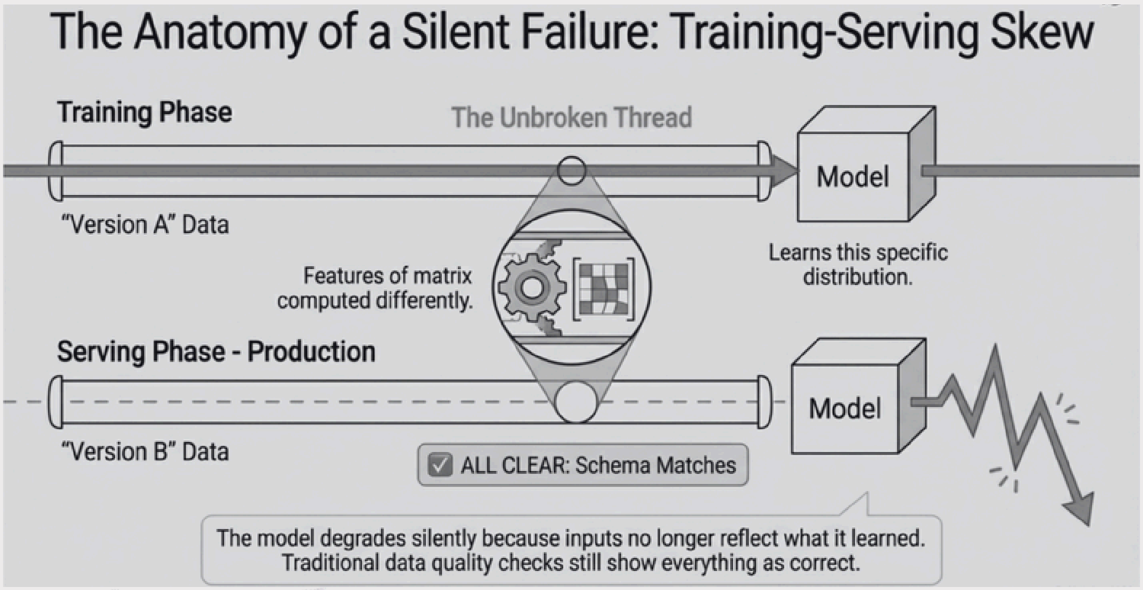
AI adoption is accelerating. Most companies are experimenting. Very few are seeing real returns. The bottleneck isn't technology—it's organizational readiness. Governance is lagging. The workforce question remains unanswered.

Agent Observability: Flying Blind at Machine Speed

Sebastian Kohlmeier
Principal PM Manager

9 min read · June 3, 2026 · Sebastian Kohlmeier

Shipping an AI agent is the easy part. Keeping it accurate, safe, and accountable in production is where teams get stuck. Agents are non-deterministic. Their behavior shifts as models update, tools change, and traffic patterns evolve and most of that drift happens silently, long after the demo. End-to-end **observability** covering the full development lifecycle is how you close that gap: See every step an agent takes, evaluate quality and safety against criteria you define, optimize what isn't working, and prove the business value of what is.



According to McKinsey, only about one-third of organisations have scaled AI from pilot to production, highlighting a clear gap: strong data foundations are not translating into reliable AI outcomes.

Understanding Ai agent observability

Agent observability is a specialized approach to monitoring AI systems that focuses on the unique behavior of agents that plan, reason, and act through tools. Traditional observability focuses on system health, including CPU, memory, error rates, and request latency. Agent observability adds semantic context: which prompt or plan the agent used, what tools it called in what order, how each step influenced the final outcome, and how the result performed against quality criteria.

Agent Observability: Flying Blind at Machine Speed

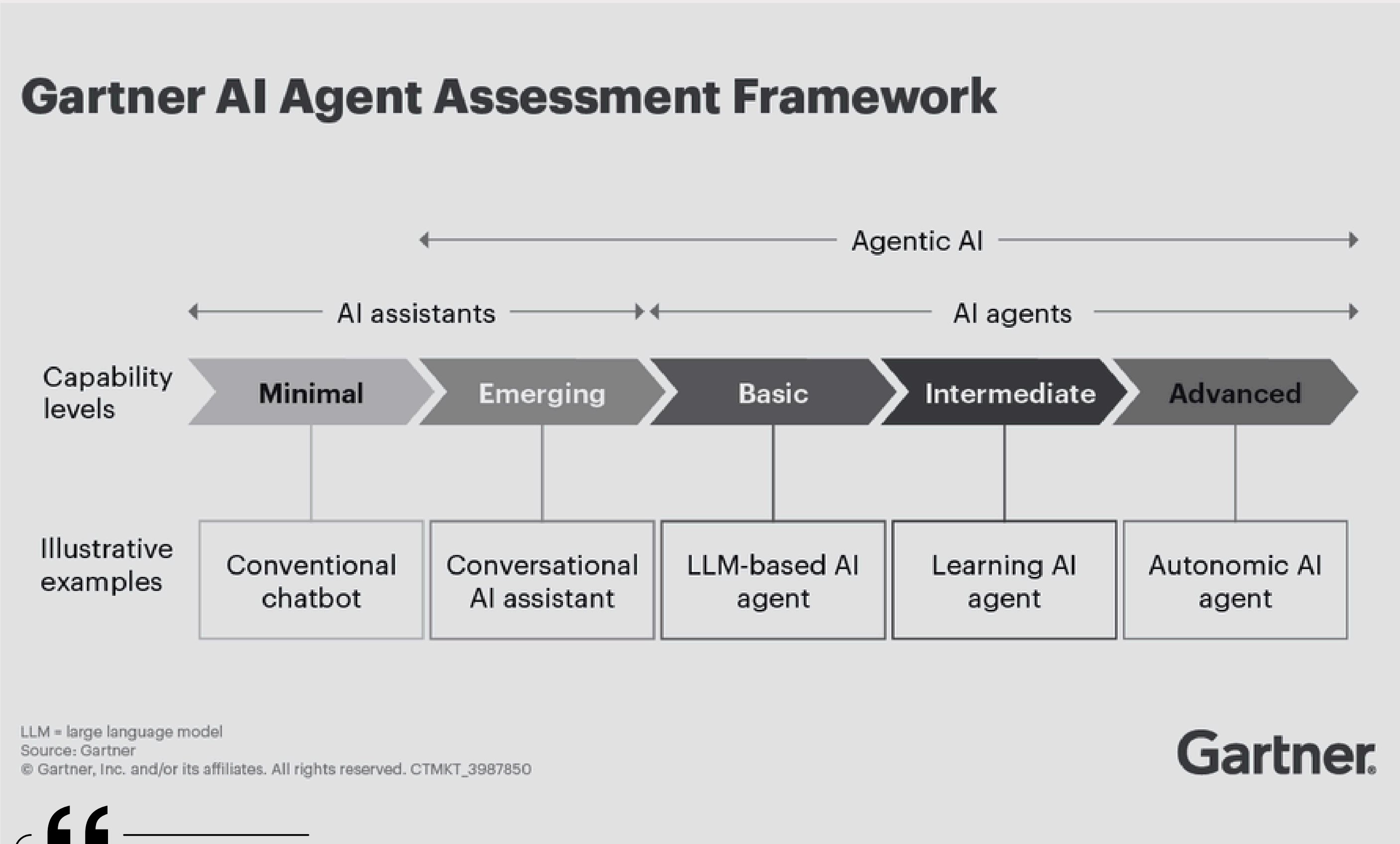
89% of teams running AI agents have observability tooling. Only 52% have evaluation frameworks.

Teams running AI agents have monitoring dashboards, latency, uptime, token usage, error rates. What they don't have is evaluation. Monitoring tells you the agent ran. It doesn't tell you if the agent was right.

Image: Summary of Gartner's AI Agent Framework | Source: Gartner

Why doesn't traditional testing work for agents?

Normal software testing assumes determinism, same input, same output. Agents break that. The same task can be completed through multiple valid paths, in different sequences, with different tool calls. You can't write a pass/fail assertion for that. Worse, an agent can produce a plausible-looking output through a completely broken reasoning path, and your dashboard won't flag it.



> By 2028, Gartner predicts that one-third of GenAI interactions will involve autonomous agents. These agents don't just respond; they act, make decisions and execute tasks independently.

Prevailing Confusions and Gaps in Observability

Observability, as it has been practised for data pipelines and software systems, measures what happened. For AI agents, the harder question is whether what happened was sound.

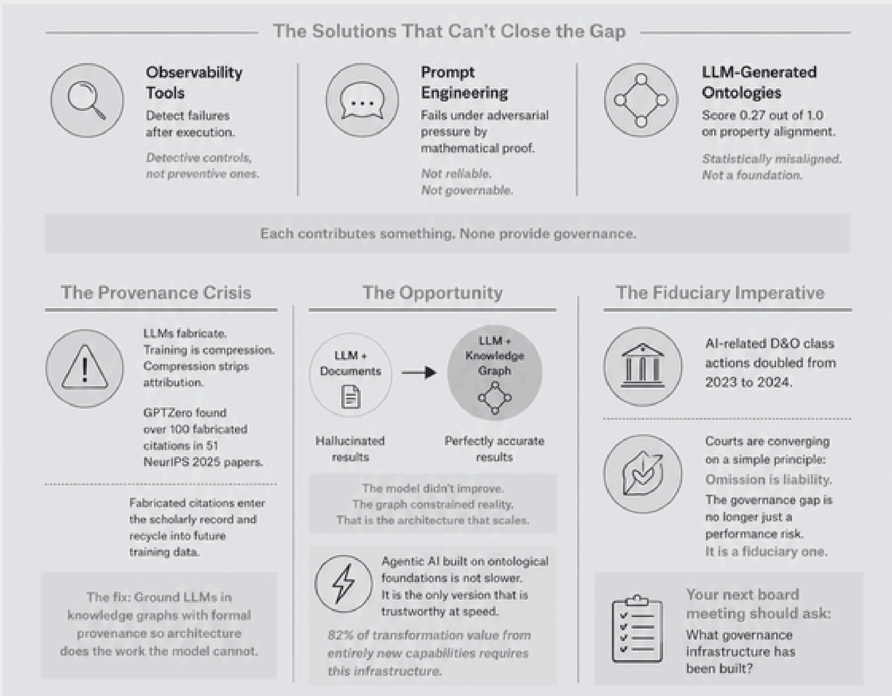
The gap is measurable. Several companies now run data quality programs, and most are extending those same observability frameworks to AI workflows, detecting anomalies, resolving data issues, checking that features and prompts meet pipeline standards. That is necessary but not sufficient. An agent can pass every pipeline quality check and still reason its way to a wrong decision through a plausible-looking path that no dashboard will flag.

Rakesh Gohel, Founder & Chief AI Strategist JUTEQ Inc, writing in May 2026, identified three observability primitives that most production teams are missing: trace IDs that follow every agent thread end-to-end, span-level token budgets enforced per tool call, and assertion layers that validate outputs semantically before they reach users. Standard logs, he noted, capture request and response pairs. They miss the reasoning in between.

Bias and lineage enter as control points that observability alone cannot close. Bias shapes how a model interprets its inputs before it acts. Lineage tracks where the data feeding that interpretation came from. Both are under-addressed even in mature governance programs, and both become compounding risks as agents are granted more autonomy to act on their outputs without a human in the loop.

The honest reckoning for 2026 is that most organisations have built observability for a deterministic world and are applying it to a non-deterministic one. The effort belongs upstream: in the semantic infrastructure, the data contracts, the governance rules that define correct reasoning before the agent runs, not the dashboards that report what happened after it did.

Image: Insights from from Frédéric Verhelst, Product Owner, Dynamics 365 & Digital Service Platforms



Prevailing Confusions and Gaps in Observability

Frédéric Verhelst, PhD · 1st
Agentic AI Governance in Critical Infrastructure | Non-Executive Director & ...
5mo ·

Observability Is Not Governance. Here's the Difference.

Capgemini surveyed hundreds of CXOs: sixty-seven percent say governance protocols are the only way to safely use AI for decisions. Only thirty-four percent have them. In that gap, vendors sell observability as a substitute. Your board thinks the problem is solved. It is not.

I believe in agentic AI. Observability is essential. But it cannot substitute for governance. The distinction matters: detective versus preventive controls.

Stuart Winter-Tear · 1st
Independent AI Advisor | AI as Capital Discipline | Author of UNHYPED | ...
View my newsletter
8mo ·

I like this. The "collaboration gap" formalises what many of us have seen: pair agents and performance often drops.

But the takeaway is philosophical, as usual. Collaboration is not an emergent property of intelligence, it's a distinct capability. And right now, it's missing. Even basic coordination under partial observability breaks current systems. That makes this a genuine roadblock to deploying general-purpose agents safely and productively.

Michał Krzyżanowski · 1st
AI & Data Transformation | Helping companies use AI to solve real business ...
Visit my website
2mo ·

From my work with enterprise clients, the gap between "working agent" and "production-ready agent" is almost never the agent logic itself. It's governance, security, and observability. This series addresses exactly that gap.

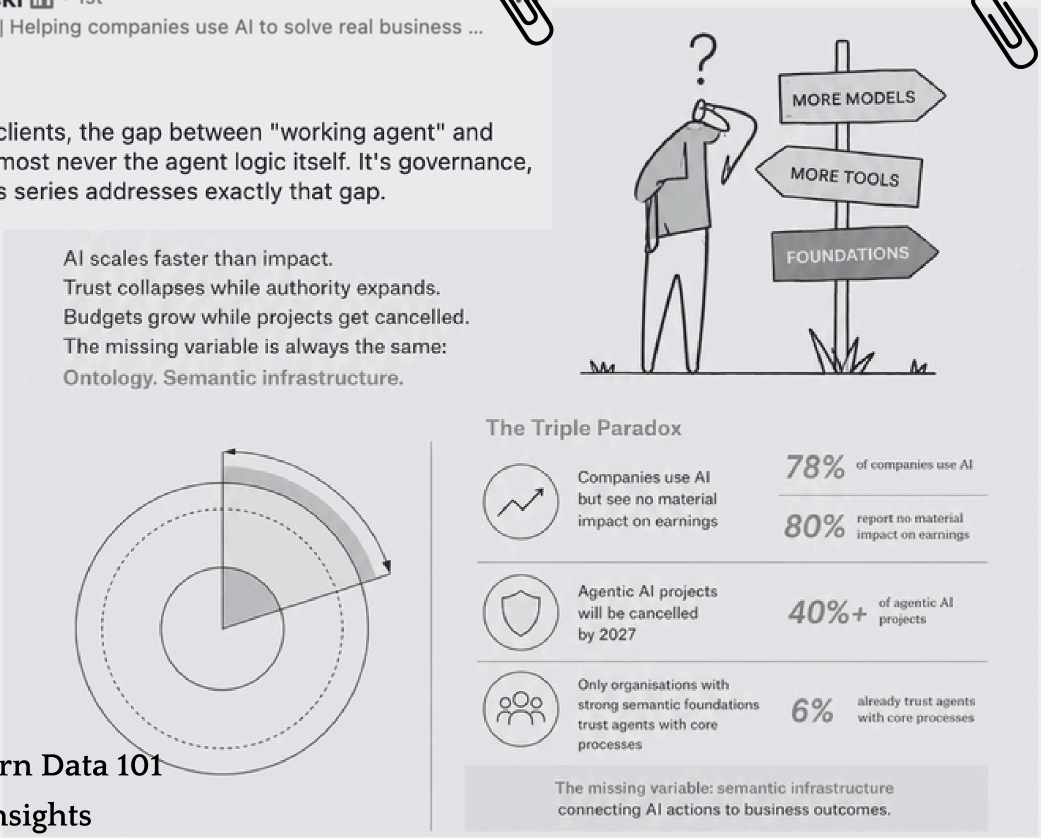


Image created by Modern Data 101 Community based on insights

The Bill Has Come Due

The question enterprises have been circling all year is not whether AI works. It is whether the way they are delivering it can survive contact with a CFO.

As one industry piece on the shift put it, Sam Altman himself has offered a striking admission that chat-model performance has already saturated and "maybe they're going to get worse." That's the pull-quote for this section, proof the downsizing conversation isn't just vendors with something to sell

”

Processing one million conversations through traditional Large Language Models costs between \$15,000 and \$75,000. The same workload through a Small Language Model? \$150 to \$800. That's not a marginal improvement — it's a 100x cost reduction that fundamentally rewrites the ROI calculus for any organization running AI at scale ~ Source

“

Why Now?

For a long time, "small model" was shorthand for a compromise, you gave up quality to save money, and the trade only made sense if the task was trivial enough not to notice. That assumption didn't erode gradually; it broke. Capability had quietly decoupled from size, and once that happened, cost became the only variable still worth debating.

Cost, it turns out, was more than ready to take center stage. Companies that rolled AI access out broadly learned the hard way that per-token pricing behaves fine at demo scale and terribly at real scale.

Instead of picking a side in the small-model-versus-large-model debate, mature teams stopped treating it as a choice at all.

Now enteprises wanted to focus on lightweight model, and reserves the expensive frontier system for the harder fraction that actually needs it.

That's a subtler idea than downsizing: it's about matching effort to difficulty rather than applying maximum firepower to every request out of habit.

Lean AI: The Downsizing & Economic Correction

The Toyota parallel driving the conversation

The most useful frame for this section doesn't come from a tech company at all, it comes from a Japanese car factory in the 1950s. Modern Data 101's Travis Thompson, Aishwarya Sharma, and Muskan Purohit trace today's "Lean AI" movement directly back to Taiichi Ohno and the Toyota Production System, where the operating question was simple: how do you produce more value with fewer resources, without sacrificing quality? Their argument is that enterprise AI has walked into the exact same trap Toyota faced seventy years ago — not a shortage of capability, but a drowning in complexity, cost, and experiments that never leave the lab.

The invoice stopped being background noise. Pilot-scale AI spending looks harmless on a dashboard. Production-scale spending does not. Enterprises that rolled broad model access out to large teams learned this the expensive way, and the resulting sticker shock is what pushed cost efficiency from a nice-to-have into a board-level metric almost overnight. Industry surveys now show a majority of infrastructure leaders naming cost reduction as their primary reason for adopting AI at all.

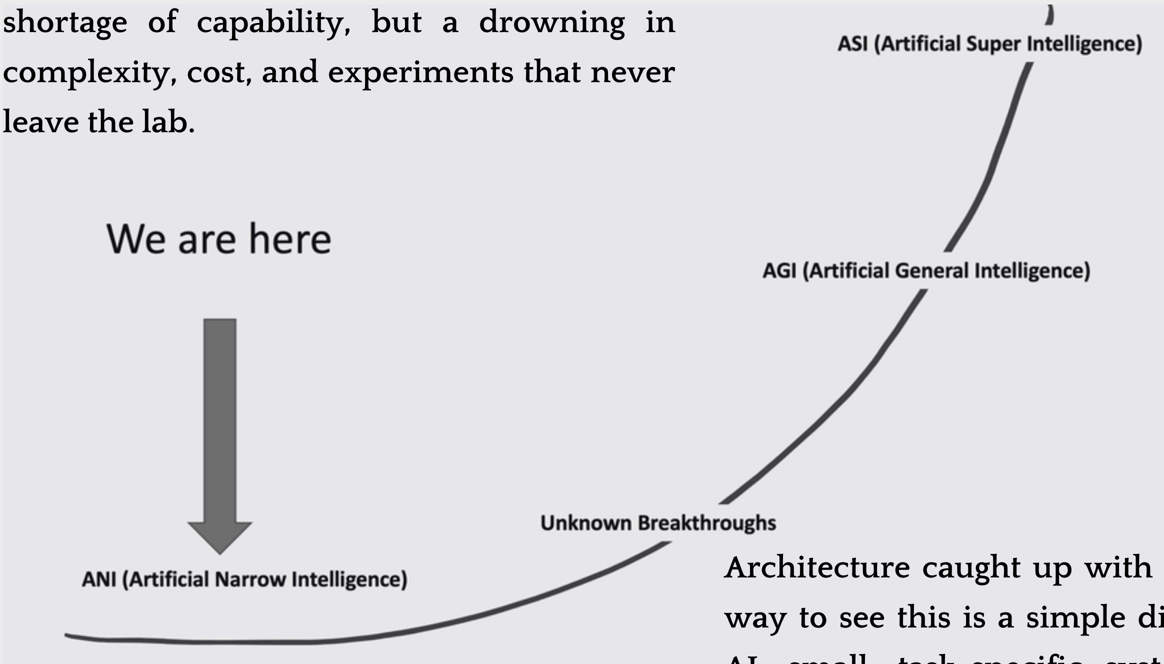


Image Source

Architecture caught up with both. The clearest way to see this is a simple distinction: Narrow AI, small, task-specific systems built into a disciplined data platform vs. Big AI, the general-purpose, frontier-model approach. Big AI isn't wrong. It's brilliant at open-ended, creative work. But defaulting to it for tasks that don't need that range is precisely the waste a lean approach exists to remove.

Illai Gescheit · 2nd
+ Follow
Entrepreneur | Investor | LBS Lecturer | Harvard Executive F...
View my newsletter
5mo ·

We need to understand that we need some sort of "Lean Planning" methods.

Planning should be more like running experiments, as a sequence of experiments that for a plan.

So when a new model, or new AI capabilities changes, you can adapt your plan.

You can adapt to the change.

Lean Enterprise Institute

Where Narrow AI Is Actually Working Now

Narrow AI and specifically large language models are what mainly drive the hype cycle and inflated claims of today. The rapid improvement in this narrower domain gives the impression we are on the edge of total AI convergence in all areas. But, in my opinion, we are not.

Görkem Sakınmaz ·
Head of Manufacturing & Operational Excellence | Lean | C...
5mo · Edited ·

If you are running an organization and diving into AI, remember that automating and digitizing a chaotic process only creates digital waste faster. Real success requires a specific triangle of discipline.

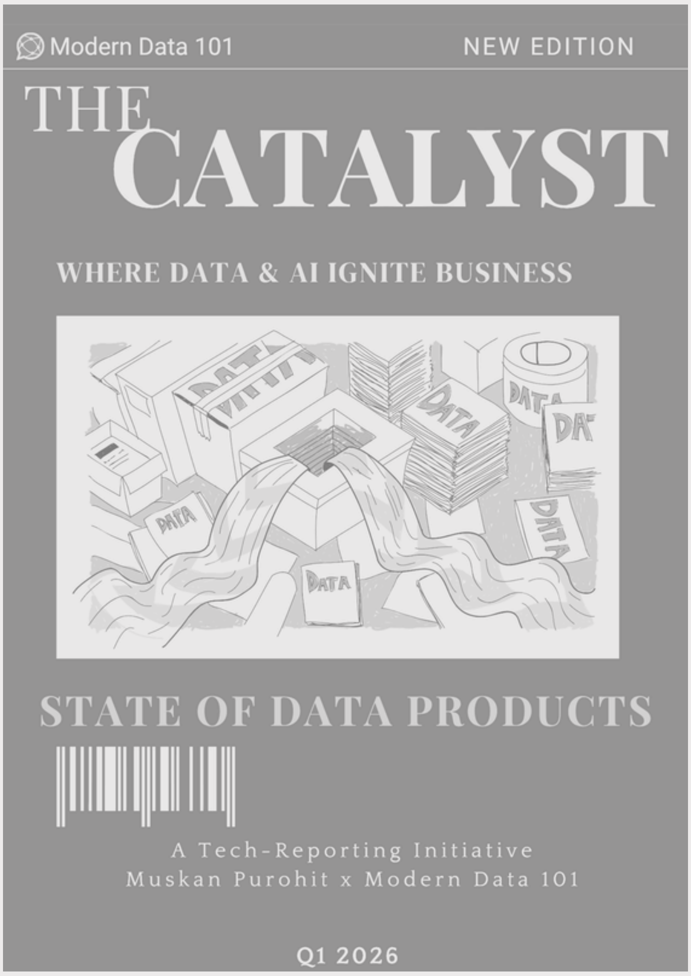
Lean Management is the foundation. You must standardize processes and eliminate waste first because AI requires clean and consistent data and process to function. Without this standardization, you are building on a swamp.

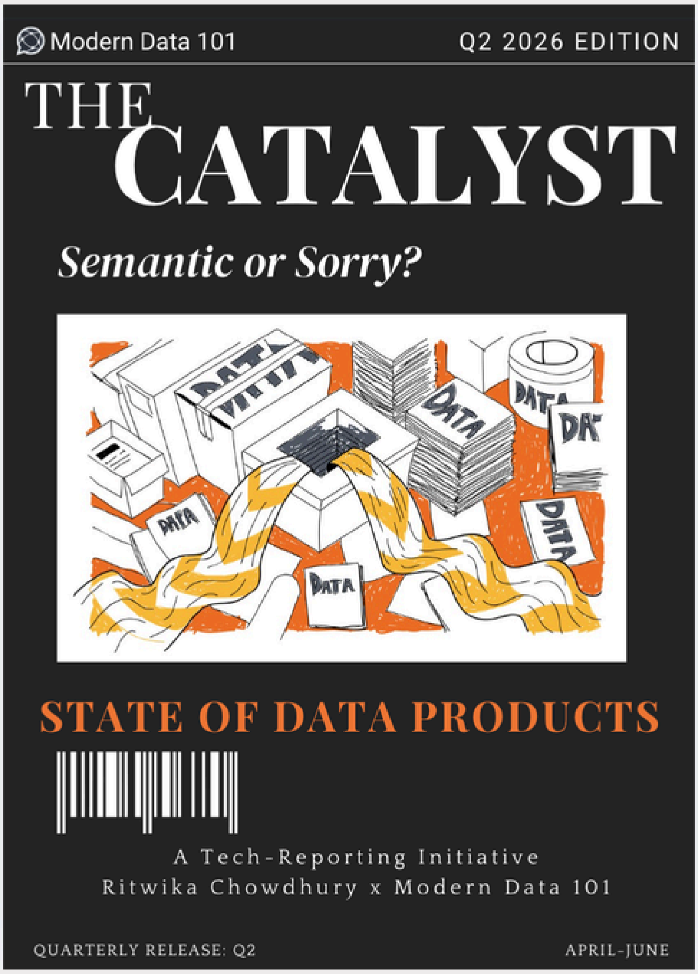
Brij Mohan
Leading all the AI/ML initiatives at TMDC with the team behind DataOS

Why Should We Think About LeanAI?

Traditional AI Framework	LeanAI Approach
Top-down roadmap	Bottom-up discovery
Vendor-driven	Domain + user-driven
Static blueprints	Continuous experimentation
Output-focused (dashboards, chatbots)	Outcome-focused (real-world impact)
No feedback loop	Fast iteration & model tuning
Centralized governance	Embedded, agile governance
Tech-first	Problem-first

Data and AI for Business
with State of Data Products





Modern Data 101

Join the Collective

moderndata101.com