# The Infrastructure Gap: Why AI Programs Are Outrunning the Systems Meant to Carry Them

*How enterprises should actually build the data and infrastructure foundation AI at scale requires*

**By Nihar Patnaik, Founder, Valtree Solutions**

## Executive Summary

*Every enterprise now has an AI strategy. Far fewer have infrastructure that can bear its weight. Deloitte's latest research found that only 42% of companies believe their technical infrastructure is highly prepared for AI — down four points year over year, meaning readiness is slipping even as AI's demands accelerate. Informatica's 2026 CDO survey found 57% of GenAI adopters still cite data reliability as a top barrier to production, essentially unchanged from the year before. IDC projects a majority of AI training data will move back on-premises within two years, reversing a decade of cloud-first assumptions. This paper argues that the gap between AI ambition and AI infrastructure isn't a technology problem — it's a sequencing problem. Enterprises are building AI capability before building the data foundation and infrastructure architecture that capability depends on. We lay out what a genuinely AI-ready foundation requires, grounded in industry research and in the patterns we've seen building this infrastructure directly, and offer a concrete starting point for closing the gap.*

## 1. The Readiness Gap Is Real, and It's Getting Worse, Not Better

There's a comfortable narrative in enterprise AI right now: pilots are proliferating, budgets are growing, and the technology is maturing quickly enough that infrastructure will catch up on its own. The data doesn't support that narrative.

Deloitte's State of AI in the Enterprise report found that the share of companies who consider their technical infrastructure strategy highly prepared for AI adoption fell to 42%, down four points from the prior year. That's not a static gap — it's a widening one. AI's infrastructure demands are compounding faster than most organizations' ability to build against them, and the perception of readiness is falling even as investment rises.

Informatica's 2026 CDO Insights survey, based on 600 chief data and analytics officers, tells a consistent story from a different angle: 57% of companies that have adopted or plan to adopt GenAI cite data reliability as a top barrier to moving from pilot to production — a number statistically unchanged from 56% the year before. This is the detail that should concern every executive sponsoring an AI initiative: the industry added an enormous amount of GenAI investment over the past year, and the core data reliability problem didn't move. Whatever effort is being spent on this isn't proportionate to the problem.

The pattern shows up again with agentic AI, which raises the stakes further because agents act autonomously on data rather than just summarizing it. Among organizations adopting agentic AI, 50% cite data quality and retrieval concerns as the top challenge to getting agents into production — ahead of security concerns at 43% and lack of agentic expertise at 42%. Data is the bottleneck before security is, and well before talent is.

The uncomfortable conclusion: infrastructure and data readiness are not catching up to AI ambition. They're falling further behind it, and agentic AI is about to make that gap more expensive.

## 2. Data Engineering Isn't a Preparatory Step — It's the Determinant of Outcome

We've built a point of view internally, drawn from doing this work inside organizations with real production stakes, that we call the Three Capabilities That Decide Every AI Outcome. The first, and most consequential, is data engineering — and the framing matters: AI does not create insight. It amplifies whatever is already present in your data. Feed it fragmented, inconsistent, incomplete data, and it will amplify fragmentation, inconsistency, and gaps with more confidence and speed than a human ever would.

We've watched organizations spend 80% of a pilot's runway firefighting broken pipelines instead of proving business value — not because the use case was flawed, but because nobody assessed whether the underlying data could support it before the clock started running. This is avoidable. It requires treating data engineering as sequenced, foundational work: source discovery across business units to find out what data actually exists and what condition it's in, ingestion and enrichment pipelines that get it into usable shape, and purpose-built feature stores designed around the specific use case — not a generic lake everyone dumps data into and nobody queries efficiently.

Encouragingly, this is starting to show up in the data. Informatica's survey found 61% of CDOs say better data quality and completeness is making it easier to move GenAI pilots into production than it was a year ago. That's real progress — but it coexists with the unchanged 57% barrier statistic above, which tells you the progress is uneven: some organizations are solving this systematically, and most are not.

**A representative pattern.** Consider a mid-size financial services company launching an AI-driven underwriting assistant. The model itself performs well in testing. But when it moves toward production, the team discovers that policy data lives across four systems with inconsistent formats, that a third of historical claims records are missing key fields, and that nobody can say with confidence which system holds the authoritative version of a customer record. The model isn't the problem. The data foundation never existed. Six months later, after building the source discovery, cleaning, and feature-store work that should have come first, the same model performs measurably better — not because it changed, but because what it was learning from finally reflected reality. This pattern repeats across industries with remarkable consistency.

## 3. Infrastructure Is Being Rebuilt Around Data Gravity, and Most Enterprises Haven't Caught Up

The second capability is infrastructure, and here the research points to a genuine strategic inflection: the cloud-first default of the last decade is reversing for AI workloads specifically.

IDC's Spotlight research on AI training and inference infrastructure, sponsored by F5, found that organizations expect 54% of the corporate data used for GenAI training and inferencing to reside on private on-premises systems within two years, versus 46% remaining in the public cloud. This is being driven by three forces: cost at AI scale, compliance requirements that are only getting stricter, and what IDC calls data gravity — the simple physical and economic reality that for very large datasets, it is cheaper and faster to bring compute to the data than to move the data to the compute.

This is happening against a backdrop of enormous scaling ambition. IDC also found that 74.4% of organizations plan to integrate GenAI into at least 11 applications, with some planning integration across more than 30 — and 75.8% already operate hybrid cloud, with 65.6% running multicloud. In other words, most enterprises are already operating in complex, distributed environments, and now need to layer AI training and inference on top of that complexity, not replace it with something simpler.

The infrastructure that supports this reliably needs four things, per IDC's framework: performance and traffic distribution at scale; application awareness — meaning the infrastructure understands the actual protocols AI workloads speak, including S3 and increasingly the Model Context Protocol (MCP); reliability across multiple environments simultaneously, since training and inference rarely happen in just one place anymore; and end-to-end security, covering encryption, strong authentication, and DDoS protection, engineered specifically against AI-native risks like data poisoning and model leakage that didn't exist as attack surfaces five years ago.

Critically, IDC's research emphasizes that none of this is achievable inside a single team's mandate. Successful GenAI deployment requires coordinated effort across network, infrastructure, storage, cloud, and data science teams that have historically operated with separate roadmaps and separate budgets. The technical challenge and the organizational challenge are the same challenge.

## 4. Governance Is What Makes AI Trustworthy Enough to Actually Deploy

The third capability — governance and architecture — is where AI programs either earn the right to scale or get stopped by their own risk and compliance functions. Deloitte's research is blunt about the underlying cause: legacy data and infrastructure architectures cannot power real-time, autonomous AI. Their recommendation is to build what they call a "living" technology and data infrastructure — modular, cloud-native, organized around domain-owned data products, with privacy, sovereignty, and security designed in from the start rather than added afterward.

This isn't an abstract architectural preference. Gartner's research on the shifting SaaS landscape makes the stakes concrete: as agentic AI rises, the priority isn't replacing systems of record like ERP and CRM — it's securely interconnecting AI agents with them, so embedded AI, custom AI, and AI-native components can coexist without fragmenting the enterprise's data architecture. An agent that can act on customer records needs an access-control and lineage model robust enough to prove, after the fact, exactly what it did and why.

The investment signal backs this up. Informatica's survey found 86% of companies plan to increase data management investment in 2026, and the top two drivers are improving data privacy and security (43%) and data readiness for AI (35%) — governance and readiness, not model capability, are where the money is actually going. The same research surfaced a complexity problem worth naming directly: companies now use an average of 7 to 9 vendor partners just to meet their data and AI management needs. That's not a governance architecture. That's a governance liability waiting to surface during an audit.

Good governance architecture — cataloging that makes data discoverable, lineage that makes outputs traceable to source, and access controls designed in rather than retrofitted — is what lets an organization say yes to scaling an AI initiative instead of stalling it in review indefinitely.

## 5. Sequencing, Not Spending, Is the Real Constraint

Pull these threads together and a clear argument emerges: the organizations succeeding with AI at scale aren't the ones spending the most. They're the ones sequencing correctly — data foundation, then infrastructure, then governance, built together rather than bolted on after a pilot proves the concept and executives start asking why it can't just go live everywhere.

We also apply a simple filter before recommending any organization fund a new AI pilot: does the data exist and is it genuinely accessible; is there a business owner who personally cares about the outcome, not just a sponsor who approved the budget; can success be measured within a defined window, typically 90 days; and if it fails, is the blast radius small — high potential value, low risk of an operational crisis. Pilots that fail these four tests don't just risk wasting budget. They risk teaching the organization, incorrectly, that AI doesn't work here — when the real lesson was that the foundation wasn't ready.

## Where to Start

1. **Run a data readiness assessment before you fund the next pilot.** Map what data actually exists across business units, its quality, and its accessibility — before committing budget to a use case the data can't support.
2. **Architect for hybrid, not cloud-only, from day one.** Given where the data is genuinely headed — IDC's research points toward a majority on-premises within two years — design training and inference infrastructure for a hybrid on-prem, cloud, and edge reality now, rather than retrofitting it once cost and compliance pressure force the issue.
3. **Build governance into the architecture, not as a compliance review after the fact.** Cataloging, lineage, and access control should be part of the initial design of any AI-facing data system, particularly as agentic AI raises the cost of getting this wrong.
4. **Get the right teams in the same room before you scale.** Network, infrastructure, storage, cloud, and data science teams need a shared roadmap, not four separate ones — this is an organizational fix as much as a technical one, and it's usually the cheapest one to make.

The enterprises that close the infrastructure gap won't be the ones with the most ambitious AI strategy. They'll be the ones who built the foundation first.