Everpure™

The Enterprise AI Guidebook

Scaling AI: The hidden challenges and how to overcome them

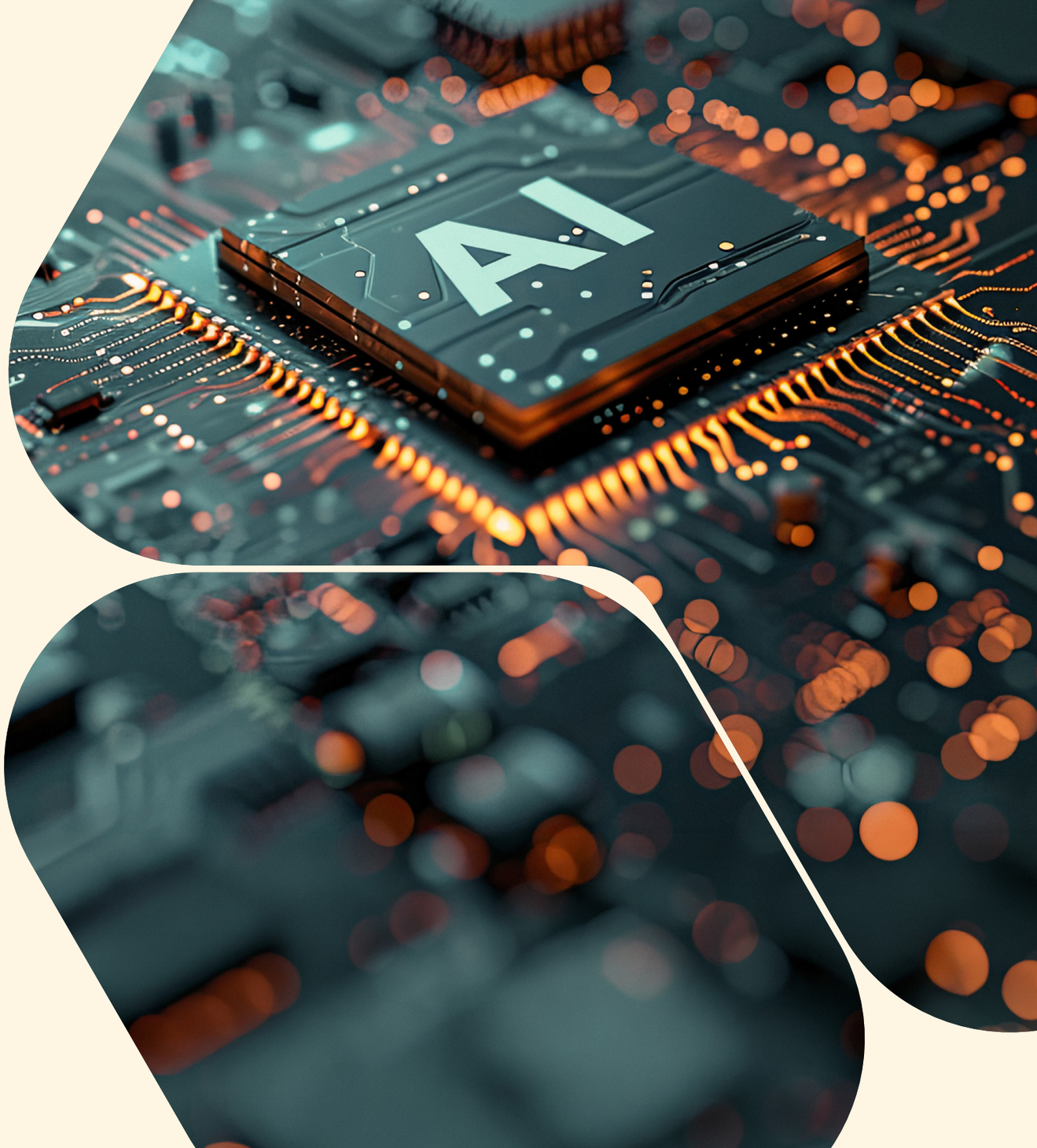

Table of contents

Introduction

AI projects start the same way in countless organizations. A business leader shows up with an idea: “Let's launch a generative AI service for customers. We need it yesterday.” Suddenly, every eye turns to IT. The pressure to deliver cutting-edge AI capabilities without breaking budgets, breaking compliance rules, or breaking the business is clear.

This is the new reality for infrastructure leaders. You’re no longer just the keeper of uptime. You’re the foundation-builder for an AI-driven future—and the pace of change leaves no room for hesitation.

The insights shared in this guide come from a discussion with infrastructure leaders who are navigating this transformation daily:

Tony Paikeday
Former Senior Director of Product Marketing, NVIDIA

Nirav Sheth
VP of Worldwide Systems Engineering, Everpure

Maciej Kranz
General Manager, Enterprise, Everpure

Wojciech Stramski
CEO, Beyond.pl

Why AI projects stall

The question isn't whether organizations want AI—clearly they do. The question is why so many initiatives stall between proof-of-concept and production. The answer lies in four converging challenges that infrastructure leaders face simultaneously.

Deployment flexibility

It's fast to get AI pilots started, but the challenge is in getting them to scale. With how quickly AI is changing, infrastructure leaders need flexibility to adapt as requirements evolve. Will your AI workload double in six months? Will it require 10 times the compute? No one knows. This uncertainty creates a planning nightmare for organizations that need to commit to capacity, budget, and architecture decisions today for workloads that may look completely different tomorrow. The winners are building flexible infrastructure that can scale up or down, move between on premises and cloud, and adapt to changing requirements without requiring a complete rebuild. Rigid infrastructure decisions made during the pilot phase become expensive constraints when AI demands suddenly shift.

Speed demands

AI is no longer a science project; it's become a business-critical, 24x7 workload. Adoption has surged from 55% of organizations in 2023 to 78% in 2024, and over 90% of Fortune 500 companies now employ generative AI technology.[1] Companies that treat AI implementation like a traditional IT project are discovering that their cautious, committee-driven approach is a competitive death sentence. While they're still evaluating vendors, their competitors are already deploying AI solutions and capturing market share.

Data sovereignty

AI workloads demand unprecedented data access. Financial services, healthcare, and government sectors need infrastructure that ensures data sovereignty, privacy, and regulatory compliance without sacrificing performance. The stakes are especially high for AI initiatives—if access to data is cut off by foreign entities or regulatory action, AI projects grind to a halt. Organizations that fail to address the issue face severe consequences. In a recent Everpure™ report on data sovereignty, 92% report potential reputational damage, while 85% cite loss of customer trust. Organizations that can guarantee sovereign, compliant, secure AI infrastructure are winning contracts competitors can't touch. This means partnering with providers who offer jurisdictional independence and compliance with local regulations—whether DORA and GDPR in Europe, DPDP in India, or APPI in Japan.

Infrastructure limitations

Legacy data pipelines buckle under the weight of AI. Infrastructure leaders cite power and grid capacity as the primary challenge—72% rate it as very or extremely challenging.[2] Supply chain disruptions compound the problem, with 65% expressing concern about component availability. Feeding massive models requires bandwidth and reliability that traditional storage simply can't deliver.

Four secrets to AI success

The organizations that thrive aren't waiting for perfect clarity; they're moving forward while others hesitate. These leaders understand that the challenges aren't roadblocks. Instead, they provide a blueprint for building competitive advantage. Here's what we can learn from these IT pioneers.

1 Stop reinventing, bring experts along

The DIY days are over. Hyperscalers might roll their own AI stacks, but most enterprises can't afford the time or cost. Smart leaders are leaning into proven partnerships, leveraging established AI architectures and turnkey infrastructure solutions to accelerate adoption without reinventing the wheel.

Don't start from scratch. Start from existing models and reference architectures. Follow established frameworks from trusted partners—like NVIDIA's AI factory model. Bring together executive leadership, infrastructure leaders, business stakeholders, and data scientists from day one. The goal isn't to build everything in-house; it's to assemble the right team and the right ecosystem of partners who've already solved the hard problems.

2 Get your data house in order

Fragmented data is the silent killer of AI initiatives. The winners consolidate data into a single data storage platform that can handle every AI workload—streamlining pipelines, eliminating silos, and making governance and security consistent across the enterprise. When data lives in disparate systems with different access protocols and management complexities, AI training becomes a constant struggle against accessibility and movement. Creating a unified data plane removes these friction points.

3 Think life cycle, not just launch

Savvy IT teams are already planning upgrade paths, exit strategies, and multicloud flexibility. They're building for adaptability, not just today's requirements. Because the only thing riskier than investing in AI today is getting locked into infrastructure that can't evolve tomorrow.

> "AI is not a 100-meter dash but a marathon."
>
> **Wojciech Stramski,** CEO, Beyond.pl

4 Build governance into infrastructure, not around it

AI pilots run on scrappy infrastructure with minimal oversight. A few data scientists, a handful of GPUs, maybe some cloud credits. Governance? That's a "production problem." Except when those pilots succeed, the governance gap becomes a production crisis. Who has access to which models? Where is training data stored and who can use it? Without governance built in from day one, scaling means retrofitting policies onto infrastructure that was never designed for them—a costly, time-consuming mess that delays production for months.

From pilot to production: Building an AI factory

Here's the blunt truth: AI that never leaves proof-of-concept is wasted investment. The gap between experimenting and scaling is where most enterprises stall. And there is a fundamental disconnect between the infrastructure you have today and what is needed to scale AI.

The limitations of legacy data centers.

Why does legacy infrastructure limit AI growth? Because the data center of 20 years ago wasn't built for AI. Traditional infrastructure was designed for storage and retrieval, not for feeding massive models that demand continuous, high-speed data pipelines. The requirements are fundamentally different: AI training requires thousands of GPUs communicating constantly, demanding high-bandwidth networking with predictable performance. Storage systems designed for occasional database queries weren't built for training runs that demand terabytes of data per hour or inference workloads hitting storage thousands of times per second. Moving from pilot to production requires rethinking infrastructure from the ground up.

This is where NVIDIA's AI factory framework becomes essential. Think of it as a manufacturing plant for intelligence. The raw material is data. The production line includes compute, storage, and networking—all optimized to work together at AI scale. The output is what your business needs: customer service responses generated in real time, questions answered instantly, and incident reports drafted automatically.

But here's the critical part.

If your data isn't ready, nothing else matters. The most powerful GPUs in the world sit idle when they can't access data fast enough. That's why NVIDIA partners with infrastructure providers that specialize in the data layer—to ensure information flows into the AI factory at the speed and scale production demands. So what does AI-ready data actually look like?

Getting your data ready

Most organizations have data. It's infused in every aspect of the business and is one of their greatest assets. What they don't have is data that's ready for AI.

AI-ready data means three things: it's accessible at speed, it's governed consistently, and it can scale without becoming a bottleneck. Legacy storage systems fail on all three counts. They were designed for occasional queries and batch processing, not for training runs that demand terabytes of data per hour or inference workloads hitting storage thousands of times per second.

A data platform built for AI

This is where partnerships with proven infrastructure providers become critical.

Companies like Everpure, with years of AI deployment experience alongside NVIDIA, provide platforms purpose-built for these demands—systems that can ingest massive data sets, feed them to compute at GPU speed, and maintain governance across the entire life cycle.

> "AI is now a 24x7 workload, and businesses increasingly need proven platforms that are enterprise-grade, with the security, reliability, performance, and scale they can depend on. The Everpure and NVIDIA partnership leverages a trusted foundation of enterprise AI success we've built over many years, for customers around the globe."
>
> **Tony Paikeday,** Former Senior Director, Product Marketing, NVIDIA

> "The data platform we're bringing to life supports the entire life cycle of AI—ingest, prep, training, deployment, and governance."
>
> **Nirav Sheth,** VP, Worldwide Systems Engineering, Everpure

Steps to data readiness

How do you get your data ready? Paired with the right data platform, taking the following steps can help you build data infrastructure ready for AI.

- **Inventory and consolidate data to ensure readiness**. Where is your data? How is it governed? How is this preserved as you bring data into modeling, inferencing, and training?
- **Modernize your data infrastructure so storage isn't the bottleneck.** Where are the different sources and how do you consolidate into a unified infrastructure and stack?
- **Build in effective AI governance from the start** with infrastructure that enforces policies automatically.
- **Partner with experts who have years of experience with AI data infrastructure**, not six months of marketing slides.

The companies that will define their industries in the next decade are making these infrastructure decisions today. The question isn't whether to modernize; it's whether you'll do it before your competitors gain an advantage you can't overcome.

The bottom line

The infrastructure decisions you make today will determine whether your organization leads or follows in the next decade. The three challenges—speed demands, governance requirements, and infrastructure limitations—aren't going away. They're intensifying. And the gap between organizations that modernize now and those that wait is widening every quarter.

> "AI is a necessity. If you want your business to be here in five years, you have to adopt."
>
> **Wojciech Stramski,** CEO, Beyond.pl

The four strategic imperatives provide your roadmap: partner strategically, get your data house in order, think about the life cycle from day one, and turn compliance into competitive advantage. These aren't theoretical concepts; they're the proven practices of organizations that have successfully moved AI from pilot to production.

AI infrastructure isn't a side project. It's a survival strategy. While competitors hesitate, the opportunity to build foundational advantage is here. The question isn't whether to act; it's how quickly you can move.

Want to learn more?

Watch the Full Conversation

1 | McKinsey & Company, "The state of AI: How organizations are rewiring to capture value," March 2025; McKinsey & Company, "AI in the workplace: A report for 2025," January 2025

2 | Deloitte, "Can US infrastructure keep up with the AI economy?" June 2025

Visit Our Website | **800.379.PURE**

Everpure™

ID18155-01-en-03/26