

THE GENAI INFRASTRUCTURE CHECKLIST

FROM PILOTS TO PRODUCTION

GenAI is a clear differentiator, but only when it runs at scale. To move from AI pilots to production, organizations need infrastructure that won't buckle under the demands of massive parallel processing.

THE ROADBLOCKS TO PRODUCTION-READY GENAI

62% of organizations reported experiencing extensive or moderate challenges moving AI models from development into production.¹ Those challenges include:

<p>1</p> <p>INFRASTRUCTURE</p> <p>Legacy systems don't have high-bandwidth, low-latency interconnects large language models (LLMs) need to function.</p>	<p>2</p> <p>TALENT</p> <p>42% of organizations lack the in-house expertise needed to build their own AI infrastructure.¹</p>	<p>3</p> <p>REGULATORY PROCESS</p> <p>The evolving regulatory landscape causes friction, slowing down deployment cycles.</p>
--	---	--

FROM LEGACY SYSTEMS TO AI-READY INFRASTRUCTURE

From	To	By the Numbers
<p>LEGACY INFRASTRUCTURE</p> <p>Hardware built for traditional workloads.</p>	<p>AI-OPTIMIZED INFRASTRUCTURE</p> <p>Systems designed to handle the extreme memory, power, and thermal demands of modern GenAI environments.</p>	<p>85% of organizations agreed they are currently incorporating their own enterprise data into generative AI solutions.¹</p>
<p>CAUTIOUS EVALUATION</p> <p>Pilots and limited experimentation.</p>	<p>ENTERPRISE DEPLOYMENT</p> <p>Scaling GenAI across teams and business functions.</p>	<p>86% of large midmarket and enterprise-sized organizations plan to invest at least \$1 million in GenAI over the next 12 months, with nearly one in five (18%) planning to commit \$5 million or more.²</p>

A GENAI TRANSFORMATION, INDUSTRY BY INDUSTRY

From healthcare to telco, retail, and beyond, we've reached a critical inflection point in today's Enterprise GenAI investment. But what does that look like in practice?

🏥 **HEALTHCARE**

<p>ADVANCING DIAGNOSTIC PRECISION</p> <p>Organizations are using GenAI for critical tasks like early cancer detection and personalized research. These models analyze massive datasets, including medical images and genomic sequences, to identify minute anomalies that enhance diagnostic speed and accuracy.</p>	<p>PROOF IN PRACTICE</p> <p>We wanted to take our existing AI capabilities to the next level by bringing in expertise from Silo AI. It is a pleasure to collaborate with Silo AI, who has extremely valuable capabilities in applying computer vision to the healthcare domain.</p> <p><i>Tero Virta, Head of R&D in Radiation Oncology, Philips</i></p>
---	---

[LEARN MORE](#) ➔

📶 **TELCO**

<p>NETWORK OPTIMIZATION AND FRAUD DETECTION</p> <p>Models analyze massive, real-time data streams to predict capacity needs, manage traffic congestion, and enhance cybersecurity measures, demanding extreme data ingestion and low-latency processing.</p>	<p>PROOF IN PRACTICE</p> <p>We are confident that AMD has the capability to fully deliver the level of performance we require.</p> <p><i>Toru Maruta, Executive Officer, Technology Planning & Strategy, KDDI</i></p>
---	--

[LEARN MORE](#) ➔

🏦 **FINANCIAL SERVICES**

<p>CREDIT RISK ANALYSIS</p> <p>Analyze large customer datasets to create advanced credit scoring models and simulate multiple scenarios, relying heavily on the ability to perform complex modeling, often using synthetic data generation.</p>	<p>PROOF IN PRACTICE</p> <p>We're trying to make technology a differentiator for our bank rather than just a necessity. Our relationship with AMD goes back almost 2 decades [from] when they introduced the first multi-core processor option back in 2005. AMD offers several capabilities, but the big benefit is value.</p> <p><i>Robert Kreitzer, Cloud Infrastructure Leader, Cloud Acceleration Program, KeyBank</i></p>
--	--

[LEARN MORE](#) ➔

🏭 **MANUFACTURING**

<p>QUALITY CONTROL</p> <p>Models analyze massive sensor and visual datasets to identify defects and predict equipment failures, requiring low-latency inference on localized, distributed compute.</p>	<p>PROOF IN PRACTICE</p> <p>Our platform can stand up authenticated, production-ready model APIs hardware. It's the full-stack approach to deployment that developers and enterprises actually need, from prototype to production in minutes.</p> <p><i>Geno Valente, Head of Go-to-Market and Engineering, LiquidMetal</i></p>
---	--

[LEARN MORE](#) ➔

🛒 **RETAIL**

<p>PERSONALIZED RECOMMENDATION ENGINES</p> <p>Utilize AI-created personalized recommendation engines that meet customers where they are. Understand your shoppers on an individual level, identifying consumption patterns, brands, and specific SKUs that they'll love.</p>	<p>PROOF IN PRACTICE</p> <p>Thanks to Amazon EC2 instances powered by AMD EPYC™ CPUs, we saved money and freed up time for our engineers to innovate on the car shopping experience for our customers.</p> <p><i>Joshua Go, VP of Infrastructure Engineering, TrueCar</i></p>
---	--

[LEARN MORE](#) ➔

A CHECKLIST FOR SCALING BEYOND AI PILOTS

Achieving true AI readiness requires progress on two fronts:

<p>ORGANIZATIONAL READINESS</p> <ul style="list-style-type: none"> <input type="checkbox"/> Commit to a GenAI-first strategy to fuel business transformation. <input type="checkbox"/> Align talent, governance, and operating models to support AI at scale. 	<p>INFRASTRUCTURE READINESS</p> <ul style="list-style-type: none"> <input type="checkbox"/> Build a resilient technical foundation optimized for efficiency and performance. <input type="checkbox"/> Adapt infrastructure to support emerging GenAI models and evolving workloads.
--	--

BUILD THE FOUNDATION FOR AI INNOVATION

Supporting the full GenAI lifecycle calls for a partner with the full-stack solution to deliver it. Scale AI without runaway costs or complexity with the full spectrum of CPUs, GPUs, accelerators, networking, and open software.

[➔ LEARN MORE](#)