



INSIGHTS | **TECH BRIEF**

## Fast-Tracking AI Implementation with HPE Private Cloud AI

Consider the headline-grabbing prediction from IDC that predicts AI will add \$19.9 trillion to the global economy by 2030.<sup>2</sup> Undoubtedly, many businesses want to be a part of it, which should come as no surprise that IDC also predicts that 80% of CIOs will embrace AI and Automation to create insight driven businesses by 2028. This surge of transformative innovation is set to propel the generative AI market from a modest \$16 billion in 2023 to a colossal \$143 billion by 2027, as enterprises race to seize their share of this technological gold rush.<sup>3</sup>

Keep in mind that these statistics are driven around intentions. Unfortunately, the best of intentions doesn't always come to fruition when it comes to Gen AI implementations:

- 90% of AI proof-of-concept pilots fail to reach near term production<sup>4</sup>
- In 2025, the Enterprises will abandon 46% of all AI pilots before they reach production in 2025<sup>5</sup>
- Gartner predicts that 30% of all Generative AI projects will be abandoned after proof of concept by the end of 2025<sup>6</sup>

### Why AI Projects Fail

Despite significant investment, most enterprise AI initiatives stumble due to a multitude of factors:

- **Technical Complexity:** Pilot-stage challenges prove more resource-intensive and time-consuming than anticipated and often require specialized expertise that many organizations lack internally.
- **Security and IP Concerns:** As projects scale, enterprises discover significant risks around data protection, intellectual property exposure, and regulatory compliance that weren't apparent during small-scale testing.
- **ROI Uncertainty:** Quantifying returns becomes difficult when pilots move to production, making it challenging to justify continued investment to stakeholders.
- **Operational Costs:** Beyond initial deployment expenses, the ongoing costs of running and maintaining AI systems in production often exceed budget projections, creating unsustainable financial burdens.

IDC predicts that 80% of CIOs will embrace AI and Automation to create insight driven businesses by 2028.<sup>1</sup>



**EP 45:** HPE Private Cloud AI - Moving from Concept to Outcomes  
*LISTEN NOW*



While well-funded enterprises may have the financial resources and internal expertise to bring their AI vision to life, the majority will struggle. An additional challenge beyond AI project completion is extracting meaningful value from it as quickly as possible. Without a clear strategy for integration, optimization, and rapid deployment, even the most advanced AI initiative risks becoming an expensive experiment rather than a transformative asset.

## An AI Turnkey Solution

Organizations have long recognized the value of turnkey solutions for critical business systems like ERP, CRM, and ecommerce platforms. This approach eliminates the complexity of custom development while accelerating time-to-value. This trend has expanded to infrastructure with services like HPE GreenLake that offers Network-as-a-Service or Storage-as-a-Service solutions. Given the technical complexity and high failure rates of generative AI projects, the logical evolution is toward turnkey AI solutions that remove implementation barriers and deliver immediate business impact.

Hewlett Packard Enterprise (HPE) Private Cloud AI (PCAI) is the turnkey, enterprise-focused AI solution that many companies have been waiting for. PCAI is a cooperative effort of HPE and NVIDIA that combines HPE ProLiant Gen12 inferencing servers, HPE AI storage, and HPE GreenLake cloud with NVIDIA's GPU computing technology and AI software stacks. This powerful combination now gives companies the pretested, fully integrated private cloud stack they need to move pilot projects to production deployments with minimal friction.

## AI Consumption-as-a-Service

Organizations looking to implement Private Cloud AI (PCAI) can choose from three flexible financing models and each one includes 24x7 reactive support, installation, and deployment.

1. **Capital Purchase:** A traditional model where the AI infrastructure is purchased outright. This model suits organizations with predictable AI workloads and sufficient capital to invest immediately.

2. **Leasing Model:** Ideal for customers with predictable capacity requirements, allowing them to spread costs over time. Leasing allows access to PCAI's scalable storage and NVIDIA NIM microservices without upfront costs.
3. **Consumption-as-a-Service (CaaS):** Best suited for organizations with dynamic or uncertain capacity needs as it enables them to scale up or down as required. This model is ideal for organizations with variable or unpredictable AI workloads, offering a cloud-like experience within a private cloud environment.

For companies that can't seem to get across the finish line with their GEN AI implementations due to budget constraints, technical complexity, or unpredictable scaling needs, HPE PCAI's CaaS model offers a lifeline. By aligning costs with actual consumption, CaaS removes the barriers of traditional AI deployments and reduces the risk of being locked into paying for unused capacity.

## Why a Private Cloud?

Public cloud platforms undoubtedly accelerate technology deployments, but turnkey private cloud solutions like HPE's PCAI deliver the same rapid deployment advantages while allowing you to maintain complete control over your AI infrastructure. With PCAI, you retain authority over hardware selection, software configuration, and resource management. Private clouds also simplify data residency challenges and regulatory compliance, keeping sensitive information within your controlled environment rather than distributed across third-party data centers. This is especially pertinent for industries such as healthcare, finance, and government.

PCAI arrives ready-to-deploy with integrated AI development tools, pre-configured model templates, and automated workflows that accelerate the transition from experimental pilots to full production environments. The AI-optimized hardware of PCAI can be delivered as a single rack to start out but can easily expand to accommodate more sophisticated applications like Retrieval Augmented Generation (RAG) configurations. All this means no more costly infrastructure overhauls to contend with as your AI requirements evolve.

## The HPE GreenLake Advantage

HPE GreenLake has established itself as a leader in turnkey Cloud-as-a-Service solutions. With PCAI, HPE has extended this proven model to address the unique demands of AI workloads. PCAI delivers the same core advantages that have made HPE GreenLake successful across other enterprise solutions:

- HPE GreenLake's management platform provides comprehensive visibility into resource consumption across storage, compute, and network components that allows for proper precise capacity planning and cost optimization.
- Pre-installed variable capacity that can automatically scale to handle sudden workload spikes to ensure AI applications never face resource constraints during critical operations.
- Enhanced incident management and support services that are specifically designed for AI workloads which is critical for AI workloads that demand high availability and reliability.



## *The Final Component*

HPE's PCAI can accelerate your advanced AI implementation from months to minutes. While the powerful HPE-NVIDIA combination provides the technical foundation, you still need a strategic partner who understands your unique business challenges, operational goals, and AI vision. You need someone who can bridge the gap between technology capabilities and business outcomes. WEI serves as that critical link, working alongside your business and technical leadership to ensure you have the right AI solution fitted for your business. Contact us and get on the proven reliable path to AI production.

### **Sources:**

1. IDC: 80% of CIOs to Embrace AI and Automation for Agility and Insights-Driven Businesses by 2028
2. AI will contribute \$19.9 trillion to the global economy through 2030, per IDC
3. IDC Forecasts Spending on GenAI Solutions Will Reach \$143B in 2027 with a 5-Year CAGR of 73.3%
4. Reasons Why Generative AI Pilots Fail To Move Into Production
5. AI project failure rates are on the rise: report | Cybersecurity Dive
6. Gartner Predicts 30% of Generative AI Projects Will Be Abandoned After Proof of Concept By End of 2025

## **About WEI**

*WEI is an innovative, full service, customer centric IT solutions provider.*

*Why WEI? Because we care. We go further.*

WEI is an expert in business technology improvement, helping clients optimize their technology environments and work efficiently. WEI works with clients to understand goals, integrate strategy with technology solutions, and leverage their current IT environment into one company-wide model to increase utilization and efficiencies around their unique business processes.