FUSION WORLDWIDE

Intelligence Report | December 2025

Agentic AI: Preparing for the Hybrid Infrastructure Surge

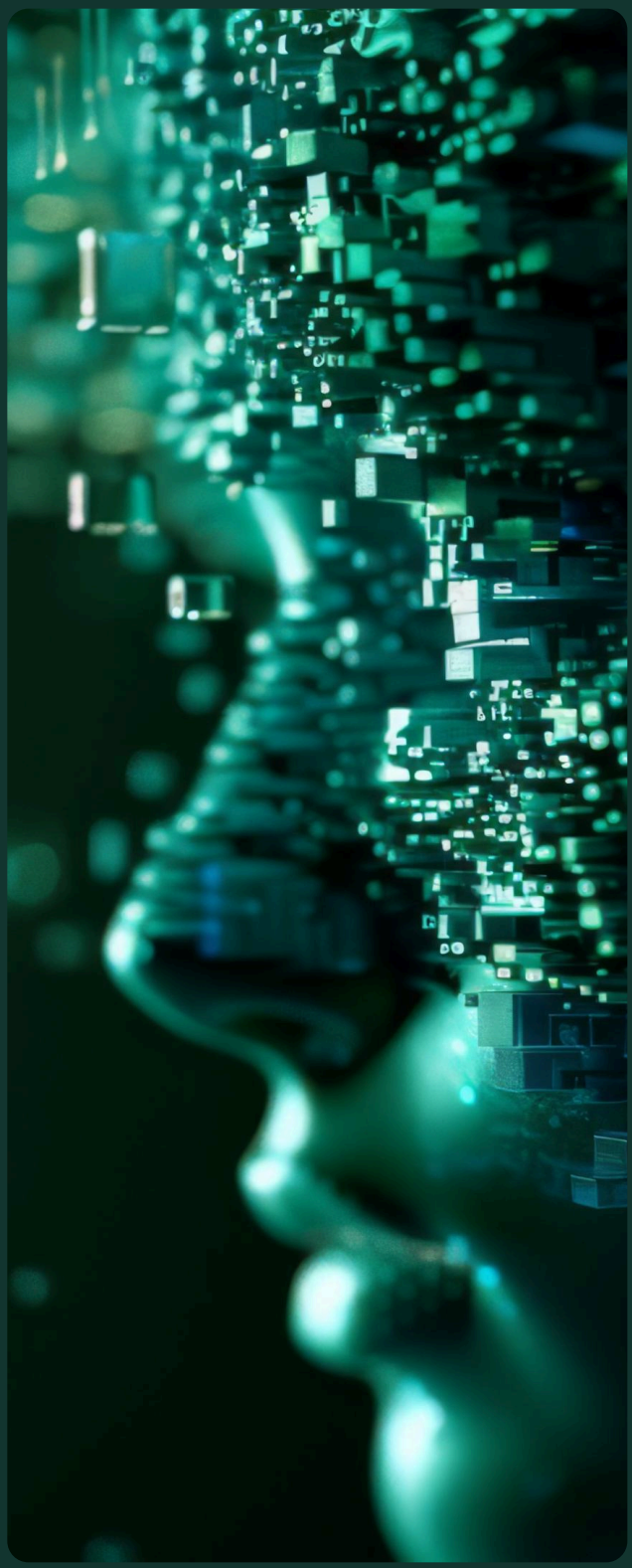
How Autonomy Will Reshape Component Supply & Demand

The State of Agentic AI

AI has reached its inflection point — intelligence no longer just responds, it acts. For years, progress was measured by what models could generate: text, images, and predictions. Now, it's measured by what they can execute. Agentic systems are beginning to manage tasks, decisions, and coordination once reserved for people, changing how enterprises think about speed, scale, and control.

The implications reach far beyond software. Agentic systems operate continuously, learn from context, and interact across connected environments. Each of these capabilities demands more computing power, greater memory capacity, and faster data movement. As deployment expands, the physical infrastructure supporting AI is scaling faster than the software itself. Intelligence is no longer dependent on computing systems — it is now shaping the systems that make it possible.

That transformation is already visible across the hardware and supply landscape. Power, memory, and connectivity have become the new levers of competitive advantage.



This report, part of the **Fusion Intelligence Report** series, analyzes how agentic AI is redefining enterprise infrastructure and accelerating the need for advanced components across the global supply chain. Drawing on Fusion Worldwide's real-time market intelligence, it identifies where demand is rising fastest, where supply pressure is intensifying, and what procurement leaders should prioritize as hybrid systems scale from pilot to production.

From Intelligence to Infrastructure: The Agentic AI Shift

Agentic AI represents a fundamental evolution in automation: systems that can plan, execute, and refine processes without human direction. Rather than generating isolated outputs, these agents coordinate workflows across applications, data sources, and physical environments.

This continuous mode of operation demands infrastructure that is always on, context-aware, and capable of real-time response. Enterprises are moving beyond cloud-only systems to hybrid models centered on on-premises mini data centers for secure, low-latency processing, while hyperscalers provide the scale for training and advanced workloads. This shift marks a new design principle for enterprise computing — intelligence as the organizing layer, infrastructure as its foundation.

For enterprise and procurement leaders, the shift isn't just technical, it's operational.
As autonomy scales, supply chains must keep pace with intelligence itself, redefining how companies plan for performance, resilience, and cost.

The Limits of Cloud-Centric Computing

Traditional cloud architectures were built for flexibility and scale, not for autonomy. For more than a decade, these first-generation systems have powered the digital enterprise, excelling at burst capacity and intermittent workloads. But continuous reasoning and 24/7 data exchange push them beyond their limits. When intelligence must act in real time, even small inefficiencies multiply quickly. Latency, data governance, and cost efficiency quickly become constraints when agents must act and adapt in real time.

Three core limitations define this challenge:

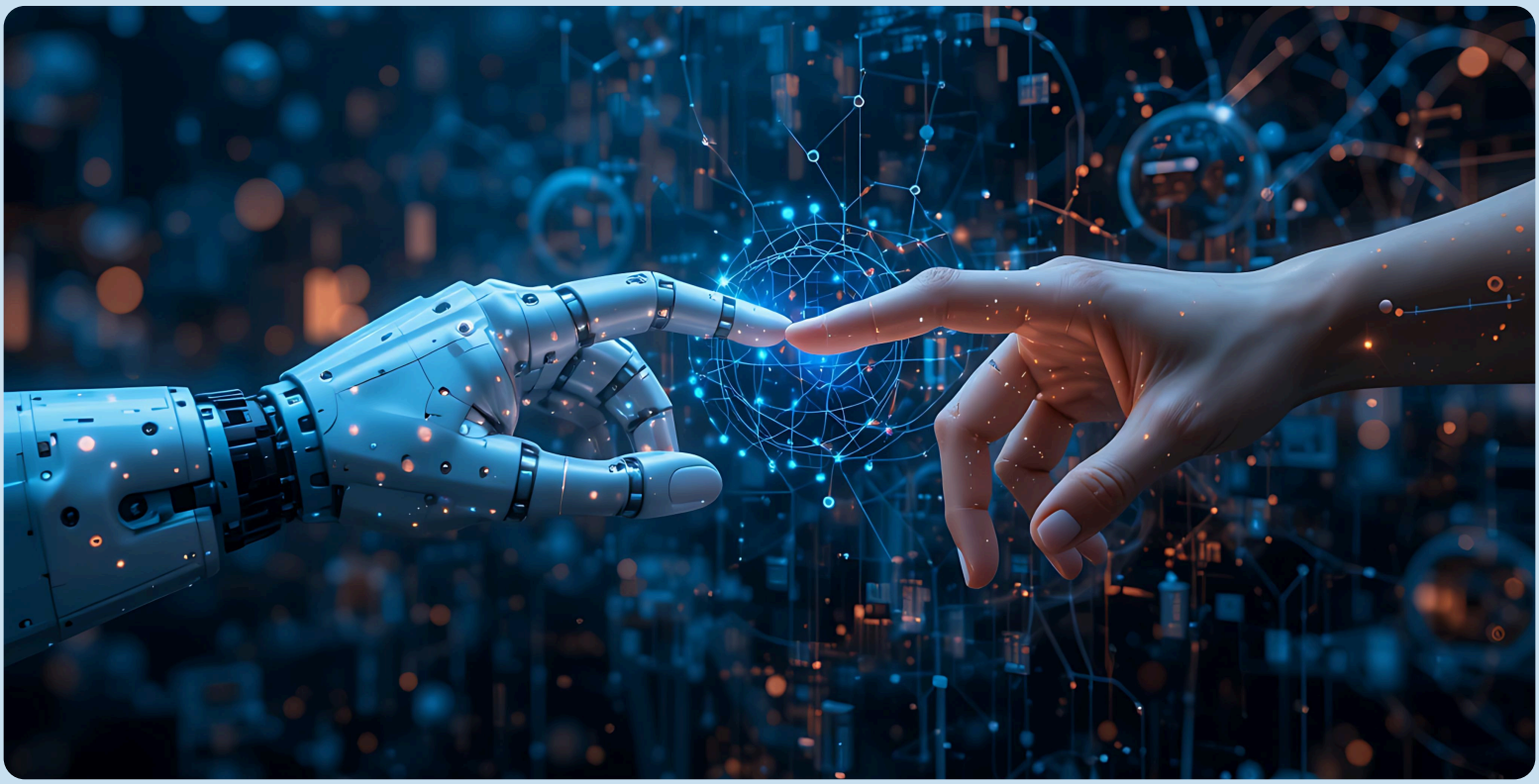
Responsiveness: Even minimal latency in cloud environments disrupts the decision cycles that agents depend on. When intelligence must act in real time, every millisecond matters

Governance: Data that fuels agentic systems is often sensitive or regulated, requiring processing within local or private environments. Centralized cloud storage alone cannot meet these compliance demands.

Sustainability: Running autonomous workloads around the clock through public cloud infrastructure drives costs and energy use far beyond sustainable thresholds.

These limits mark a structural shift in computing. As autonomy becomes the default mode of operation, enterprises are redistributing compute power closer to where intelligence operates. Edge and on-premises systems now manage real-time workloads while the cloud supports training and coordination. The result is an architecture built for speed, control, and continuous intelligence.

For enterprise leaders, this represents the practical turning point: the move from cloud convenience to compute control. The next step is understanding how organizations are adapting their infrastructure strategies to keep up.



How Organizations Are Responding

As agentic AI becomes operational rather than experimental, the weaknesses of cloud-only computing are becoming impossible to ignore.

To overcome these limits, companies are rethinking infrastructure design. Hybrid computing environments are emerging as the new standard — combining the control and security of on-premises systems with the scale and flexibility of the cloud.

More than 80 percent of global enterprises plan to integrate agentic AI within the next three years, with most prioritizing hybrid strategies that balance performance, governance, and cost efficiency. What began as isolated deployment is evolving into enterprise-wide infrastructure planning, redefining how organizations build for scale and continuity.

This evolution marks the transition from experimentation to execution. We're at the point where hybrid design becomes the operational core of autonomy.

Reengineering the Foundations for the Age of Autonomy

As agentic AI moves from research to deployment, enterprise infrastructure is being rebuilt for continuous operation. Organizations are designing hybrid systems that balance data sovereignty, latency, and cost efficiency, allowing AI agents to function autonomously across both digital and physical environments.

Agentic systems represent a new phase of automation — one defined by independence and interconnection. They plan, execute, and refine processes in real time, coordinating workflows across applications, data sources, and physical assets. Supporting this autonomy requires reliable, high-performance computing. Enterprises are therefore constructing hybrid environments built around localized data centers for secure, low-latency processing, while hyperscalers provide scale for training and specialized workloads.

This transition is creating deep shifts across the technology supply chain. Continuous operation and distributed compute are driving sustained demand for advanced semiconductors, high-bandwidth memory, and edge components. The market is now entering a cycle defined by broader component demand, regional diversification, and sovereign infrastructure initiatives — extending far beyond the temporary shortages of 2021–2022. Based on **Fusion Worldwide's market intelligence,** the next 12 months will bring structural changes that redefine where and how demand materializes across the supply chain.

What to Expect Over the Next 12 Months

- **Broader Demand Resurgence:** Growth extending beyond GPUs to include HBM, liquid cooling systems, 800G/1.6T optics, and ABF substrates.
- **Diverging Price Trends:** Premiums for top-tier accelerators and HBM4, while legacy components such as DDR4, SATA SSDs, and air-cooled systems decline in relevance.
- **Regional Fragmentation:** The US, EU, and Asia-Pacific are advancing sovereign mini data center programs in response to data localization mandates.
- **Shifting Bottlenecks:** Constraints are moving from silicon supply to power availability, skilled labor, rare-earth materials, and advanced packaging.

The Emerging AI Infrastructure Stack

Agentic AI is reshaping the foundation of enterprise computing. To support continuous reasoning, fast memory access, and localized decision-making, organizations are building hybrid models that combine edge, on-premises, and cloud environments.

The agentic AI stack is taking shape across three layers of deployment:

- **On-Premises Mini Data Centers:** Growth extending beyond GPUs to include HBM, liquid cooling systems, 800G/1.6T optics, and ABF substrates.
- **Hyperscaler Burst Capacity:** Provides training scale and peak compute flexibility.
- **Orchestration Software:** Connects and governs distributed systems, managing security, routing, and workload allocation.

This emerging architecture marks a shift from centralized processing to distributed intelligence. Each layer introduces new requirements for compute, memory, and power efficiency, reshaping priorities across design, sourcing, and deployment.

Inside the Agentic AI Hardware Ecosystem

The shift to agentic AI begins at the component level. As workloads become continuous and distributed, the entire hardware ecosystem — from compute to cooling — is being reengineered for speed, efficiency, and endurance.

Compute: The Intelligence Engine

Processing has moved from large centralized GPUs to a mix of specialized chips.

- **GPUs Remain Foundational:** NVIDIA's Blackwell remains dominant, but inference is shifting to distributed environments.
- **Custom Silicon Accelerates:** Hyperscalers invest in purpose-built AI accelerators, driving chip revenue toward $90B by 2027.
- **CPUs Regain Strategic Importance:** Modern CPUs orchestrate data flow across accelerators and networks.
- **Edge Processing Expands:** Compact AI SoCs handle local inference, reducing data transfer demands.

Memory: The New Bottleneck

As agents process continuous data, memory is becoming the limiting factor.

- **HBM Dominates Growth:** Demand projected to rise 400% by 2027, with capacity sold out through 2026.
- **DDR5 Powers the Edge:** High performance in compact, power-efficient designs.
- **Persistent and Hybrid Memory:** Ensures continuity across sessions.
- **Optimized Hierarchies:** Tiered memory architectures balance latency and cost.

Data Infrastructure: Connecting, Caching, & Powering Scale

As agents process continuous data, memory is becoming the limiting factor.

- **Storage Convergence:** NVMe, PCIe 6.0, and tiered caching enable high-speed access.
- **Networking Expansion:** 400G–800G Ethernet, 5G/6G integration, and edge connectivity support real-time coordination.
- **Power and Cooling Demands:** Data centers now consume up to 50x more energy per rack; liquid cooling adoption is accelerating 1,000%.
- **Power Management ICs:** Critical for efficiency, safety, and reliability under continuous load.

Industry Infrastructure Strategies: From Experimentation to Execution

Enterprises across sectors are redesigning their computing environments to support agentic AI, but each industry's path reflects its own operational, regulatory, and performance realities. Together, they illustrate how autonomy is becoming embedded in the physical fabric of business.

Manufacturing: Autonomy Meets the Factory Floor

Manufacturers are leading the transition with on-premises systems enabling closed-loop automation and predictive maintenance. Edge deployments allow AI agents to manage operations in real time, driving demand for ruggedized compute, industrial networking, and efficient inference chips.

Healthcare: Balancing Privacy & Precision

Healthcare providers are adopting hybrid architectures that keep sensitive data on-premises while connecting to cloud-based model training. Compliance with privacy regulations is accelerating investment in secure storage, encrypted networking, and verifiable orchestration systems.

Automotive: Intelligence at the Edge

Automakers and mobility companies are deploying agentic systems for simulation, routing, and real-time vehicle coordination. These architectures combine in-vehicle, edge, and cloud computing, fueling demand for embedded AI processors, high-speed optics, and hybrid connectivity solutions.

Finance: Contained Autonomy for Compliance & Risk

Financial institutions are building private AI environments to maintain control over data lineage and decision transparency. These systems rely on high-bandwidth memory, encryption accelerators, and low-latency compute clusters to operate within strict governance frameworks.

Across industries, AI is evolving from a discrete application to the operational backbone of enterprise systems. This momentum is setting a new tempo for infrastructure investment, one that will define the next five years of deployment and scale.

Timeline & Adoption Curve

As adoption expands across industries, deployment timelines are accelerating. Early pilots are evolving into production environments, and the next five years will define how quickly hybrid agentic infrastructure scales from innovation to standard practice.

Early Adopter Phase — 2025–2026

Pilot projects validate performance & establish early ROI benchmarks.

- Tech-forward enterprises are deploying initial mini data centers for specific use cases.
- Component availability remains constrained, with HBM and liquid cooling facing long lead times.
- Pricing is elevated but beginning to stabilize as production ramps.
- Early adopters are establishing best practices and validating ROI models.

Mainstream Adoption — 2027–2028

Enterprise-wide rollouts drive scale, standardization, & supply challenges.

- Over 50% of enterprises will shift workloads to local data centers by 2027.
- Supply chain stress points will peak as demand surges.
- Component pricing will moderate as production scales.
- Standardization will begin to emerge, reducing deployment complexity.

Maturity & Optimization — 2029–2030

Hybrid systems reach equilibrium as efficiency & cost take priority.

- Second-generation architectures will deliver improved performance and reliability.
- Component supply will align more closely with demand.
- Focus will shift from deployment to optimization and cost reduction.
- New use cases will emerge as agentic infrastructure becomes ubiquitous.

Supply Chain Implications

Agentic AI is changing how companies source and manage components. Procurement is shifting from large, centralized orders to distributed purchasing models with longer planning cycles. Lead times are extending, and advanced planning has become essential.

From Centralized to Distributed Procurement: Enterprises are moving from bulk hyperscale orders to hundreds of smaller regional buys, requiring tighter coordination and visibility.

Longer Lead Times: HBM, liquid cooling systems, and high-speed optics now carry 34–52 week lead times; procurement teams must plan 12–18 months ahead.

Inventory Strategy Reversal: Companies that reduced stock in 2024–2025 are rebuilding safety inventory for AI-critical parts as just-in-time models give way to just-in-case planning.

What Comes Next:
Competing in the Agentic Era

The next phase of AI will be shaped more by hardware availability than model innovation. Through 2030, demand for advanced compute, HBM, and power systems is expected to exceed global capacity. Procurement teams that move early by expanding supplier networks and securing long-term agreements will be best positioned to manage volatility and cost.

Traditional procurement cycles cannot keep pace. Lead times already stretch past a year, and regional production constraints are tightening. Static sourcing models cannot adapt fast enough. The leaders will treat supply networks as dynamic systems that anticipate disruption and make decisions with the speed of the intelligence they support.

Next Steps for Procurement Leaders

Anticipate constraints before they materialize. Model supply & capacity risks 12 to 18 months ahead, especially for HBM, advanced substrates, and power systems.

Diversify with intent. Expand sourcing across regions, technologies, & supplier tiers to reduce exposure to single points of failure.

Design for convergence. Align components and architectures for hybrid environments that span on-premises & hyperscale systems.

Strengthen visibility. Partner with suppliers capable of providing real-time market intelligence & proactive allocation strategies.

At Fusion Worldwide, the rise of agentic AI isn't a forecast. It's already reshaping procurement. The leaders won't be those who react to shortages, but those who anticipate them. Resilience will come not from scale, but from speed: the ability to spot shifts across the supply chain and act before anyone else. **The future of agentic AI will be built by those who move first. The time to get Out in Front™ is now.**

Contact Fusion Worldwide

For questions about how **rising AI demand** may impact component availability or sourcing timelines, Contact Fusion Worldwide or visit fusionww.com

Americas

Global Headquarters
Boston
+1 617 502 4100
boston@fusionww.com

EMEA

Regional Headquarters
Amsterdam
+31 20 667 6020
amsterdamoffice@fusionww.com

APAC

Regional Headquarters
Singapore
+65 6311 5250
singaporeoffice@fusionww.com