MIC
CEO VISION
February 2026

Why AI Services Are Redefining Semiconductor Competition

Peter Chou

Senior Industry Analyst

Peter Chou's core research focuses on compound semiconductor industry and market trends, including the impact of technology and industry changes on the global compound semiconductors, competition and cooperation relationships among major semiconductor brands, emerging semiconductor technologies, and applications of emerging compound semiconductors. He participates or has participated in several government research projects commissioned by the Ministry of Economic Affairs (MOEA) such as "Electronic Equipment Industry Promotion Program", and National Science and Technology Council (NSTC) "Next-Generation Compound Semiconductor Project". Peter holds a Master's degree from the Institute of Electronic Engineering, National Taiwan University of Science and Technology.

The Shift in Semiconductor Competition

In the AI era, semiconductor competitiveness is no longer defined by single-chip performance alone. As AI workloads scale in size, duration, and complexity, competition is increasingly shaped by system-level capability—specifically, the ability to sustain reliable and efficient operation at scale.

This shift is most evident in large-scale AI workloads such as large language models (LLMs), which depend on the coordinated operation of vast numbers of computing nodes. As model sizes and data volumes continue to expand, incremental gains in chip efficiency are no longer sufficient to drive overall AI performance. Instead, outcomes are determined by how effectively the entire computing system functions as an integrated environment.

In this context, memory access efficiency, interconnect architecture, power delivery, and thermal management have become deeply interdependent. Performance is constrained by the weakest link in the system rather than

by peak capability at the component level. As a result, AI compute capability is increasingly evaluated by whether a system can operate stably and continuously under real-world conditions.

Accordingly, the semiconductor industry is moving away from pursuing the performance limits of individual chips. The strategic challenge is no longer simply sustaining chip-level performance growth, but redefining the industry's role within an AI ecosystem that prioritizes scalability, reliability, and long-term operability.

System Optimization as the Core

As AI computing scales, AI data centers have evolved into complex systems in which performance is determined by how effectively power delivery, interconnects, computing architectures, and software frameworks operate together. When any single layer becomes a bottleneck, improvements to individual components deliver diminishing returns.

Under these conditions, competitive differentiation is shifting toward holistic system optimization across power, communication, computing, and software architectures. The objective is no longer to maximize isolated performance metrics, but to ensure that AI systems can operate stably, scale predictably, and be deployed repeatedly under practical constraints of power, cost, and deployment.

This transition is reflected in the strategic evolution of leading players such as NVIDIA, whose focus has expanded from GPUs to data center–level system design. Beyond advancing computing and interconnect technologies, it has actively shaped requirements for power distribution, data transmission, and emerging communication solutions such as silicon photonics to ensure predictable system performance under high computational density.

As data centers become the primary deployment environment for AI workloads, performance evaluation increasingly incorporates power efficiency, communication stability, and operational reliability alongside compute capability. System optimization is no longer a supporting technical function; it has become a foundational capability for long-term competitiveness in high-end AI computing markets.

Implications for Semiconductor Companies

From Chip Performance to AI Service Enablement

As AI applications shift from training to deployment, competitiveness is no longer defined by peak computing performance. Instead, it depends on whether AI systems can operate reliably over time, be maintained efficiently, and tolerate failure during service delivery. In this phase, chips function as system-embedded components rather than standalone products.

If semiconductor companies continue to compete primarily on single-chip performance, they risk being confined to limited segments of the value chain and failing to meet the operational demands of AI services. The strategic priority is to reposition chips as building blocks designed for long-term service operation.

To do so, semiconductor companies must leverage their strengths in process technology, packaging, testing, and system integration—capabilities that directly support operational stability. Closer collaboration with system providers and end users is essential to align chip design with real deployment conditions, reduce system integration barriers, and enable rapid replication at scale.

This shift moves semiconductor companies upstream in the AI value chain. Success will depend less on maximizing single-

point performance and more on enabling AI services to run predictably and sustainably in real-world environments.

From Chips to AI Services

Under an AI service-oriented development model, the semiconductor industry's role is increasingly defined by its ability to enable scalable and reliable AI service operation. While high-end chip performance remains important, design priorities are now shaped by whether systems can sustain stable operation throughout the service lifecycle.

For Taiwanese semiconductor companies, advancing industry positioning does not require abandoning high-end AI chip development. Rather, it requires reframing chip design around the operational requirements of AI services. By treating long-term system stability as the primary objective, these companies can develop deployable, maintainable, and scalable solutions that integrate seamlessly into AI service environments—even without direct participation in AI model or platform development.

As competition extends from components to systems and services, semiconductor companies must move beyond cost and specification comparisons and build deeper, long-term relationships with system integrators and cloud service providers. Competitive advantage will increasingly be determined by whether companies can reposition themselves as essential enablers of AI service operation.

From a systems perspective, the ability to support stable, long-term AI services will define who secures a durable position within the AI service ecosystem.