WHITE PAPER



Industry Insights: End-to-End AI Validation for 1.6T Networks

Expert Perspectives on AI Fabric Challenges from Physical Layer Testing to AI Deployment Readiness

KEYSIGHT

The New Validation Reality for AI Infrastructure

From switching silicon and optical interconnects to large-scale AI fabrics, industry experts point to a common conclusion: AI infrastructure is redefining what engineering teams must validate before a network is truly deployment ready.

As AI data center network speeds scale from 400G to 800G to 1.6T, tighter design margins and increasing interoperability complexity are changing where risks appear. Issues once contained in the lab are now more likely to surface in deployment. For teams across the AI infrastructure ecosystem — from chip designers and system architects to network engineers and hyperscale operators — the challenge is no longer validating individual components or links in isolation. The focus is shifting toward understanding how the complete infrastructure behaves under real operating conditions.

This shift extends beyond faster lanes and denser optics. AI workloads generate sustained east-west traffic across GPU clusters, making network performance more sensitive to latency, congestion, and workload imbalance. These conditions can expose behaviors that traditional validation approaches were not designed to capture.

As a result, validation requirements are evolving. Compliance and interoperability remain essential, but readiness increasingly depends on whether the full solution can be stressed and qualified under realistic AI workloads and deployment conditions.

This change fundamentally redefines what it means for AI infrastructure to be deployment ready. End-to-end validation now means moving beyond isolated component-level checks to evaluate how infrastructure performs at scale under realistic workloads. That is where traditional validation reaches its limits and a broader system-level approach to AI infrastructure readiness becomes essential.

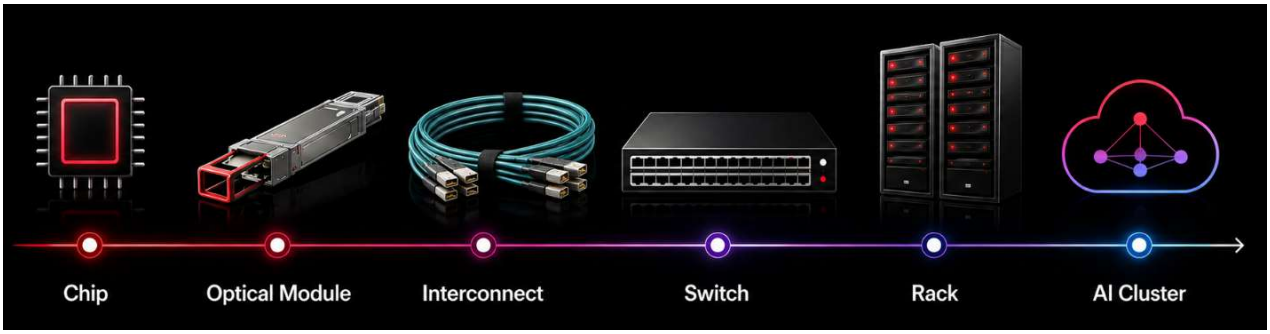


Figure 1. End-to-end validation across the AI infrastructure stack, from chip to AI cluster

To better understand this shift, experts from across the AI infrastructure ecosystem shared perspectives on how validation requirements are changing. Their insights reveal several common themes that point to a broader change in how engineering teams must evaluate readiness. The following sections examine these themes in greater detail and show how they are reshaping validation across the AI infrastructure.

What industry experts reveal:

- Higher speeds reduce margins, increasing sensitivity to errors and link instability
- Lab validation alone cannot fully predict deployment behavior
- Interoperability spans optical modules, switches, interconnects and GPU clusters
- AI-scale deployments require realistic emulation and automation

Why Traditional Validation Reaches Its Limits at 1.6T

Validating 1.6T network components is increasingly complicated due to reduced margins, increased interconnectedness, and unique AI workload traffic patterns. Each factor creates validation pressure, but together they change how even small failures can emerge and propagate throughout the entire system.

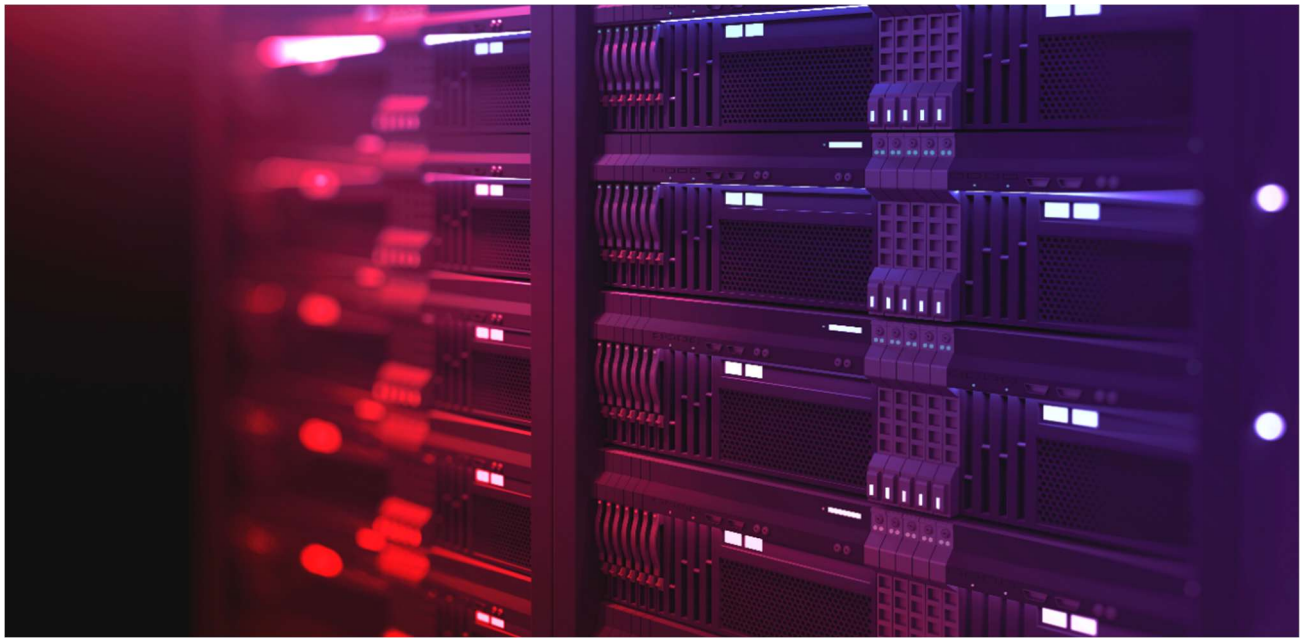
At the physical layer, the move to 224 Gb/s serializers / deserializers (SerDes) reduces available margin in electrical and optical lanes. Systems can no longer tolerate marginal behavior in connectors and cables as in previous generations. Issues that once have remained hidden are now more likely to surface under load. This shift demands higher precision and signal integrity to reliably and repeatedly characterize lane-by-lane performance at the margin.

> *"We may not provide actual margin to [the] system because, [with] 100G, we may have [a] safety buffer. However, [at higher lane rates], everything [becomes] more challenging, and we lose that safety buffer, we cannot build out those bad designs."*
>
> **Cathy Liu, Distinguished Engineer, Broadcom**

Reduced margins become more consequential in highly interconnected AI data center environments. In large clusters running synchronized workloads, congestion, latency variation, and interconnect instability are more likely to surface during operation as traffic moves across the fabric. These issues can influence job completion time, GPU utilization, and overall infrastructure efficiency, turning what appears minor in isolation into a system-level constraint at scale.

> *"[At 1.6T], I think you see a lot more issues happen in the field while the clusters are operational."*
>
> **Rakesh Cheerla, VP of Products at Netpreme Corp**

These operating dynamics are difficult to reproduce with simplified or synthetic testing approaches. As a result, the gap between lab validation and production behavior becomes more pronounced. AI fabrics that appear stable during lab validation may behave differently when fully populated, exposed to real traffic, and operating under sustained load, thermal variation, and other deployment stressors.



When Lab Validation Does Not Predict Deployment Behavior

Industry perspectives on 1.6T networks reveal a broader validation experience in which isolated checks do not always reflect how AI infrastructure behaves after deployment.

Compliance and component-level validation remain necessary, but they are no longer sufficient. A component can meet specifications yet still contribute to performance or reliability issues when integrated into a full system. Similarly, networks can pass baseline requirements while failing to deliver predictable performance under real workloads.

Interoperability is no longer a point-to-point exercise. It is a system-level challenge shaped by interactions between optics, interconnects, switches, deployment configurations, congestion behavior, and workload synchronization. As the number of variables increases, teams cannot practically test every configuration under every operating condition. Instead, validation must focus on emulating the real-world conditions most likely to expose risk before deployment.

> *"We do not limit ourselves to qualify the single element. We are really building the entire end-to-end solution from the optical module to the platform to the traffic generation [and] we always run traffic on every single test."*
>
> **Alessio Ferrari, Senior Cloud Optical Network Engineer, Microsoft**

Hyperscale operators are increasingly responsible for integrating and validating broader portions of the infrastructure themselves. This further increases the need for realistic, system-level testing earlier in the design and development cycle.

Another limitation is the reliance on synthetic or simplified test inputs. AI clusters exhibit burst behavior, electrical noise, and jitter patterns that simple pseudo-random bit sequence (PRBS) testing fails to replicate. The point is not to replace conventional methods entirely, but to extend validation toward conditions that better reflect how AI systems behave in operation.

> *"We need to test with the real AI workload instead of only test[ing] on fake PRBS data. In AI data centers, these clusters and the huge [burst of] data will create some electrical noise and jitter that simple test[s] miss."*
>
> **Cathy Liu, Distinguished Engineer, Broadcom**

Infrastructure may appear stable under simplified tests and yet behave differently under real operating conditions. In AI data centers, tightly synchronized fabrics depend on consistent performance, where even small inefficiencies, such as added latency on a single link, can affect performance at scale.

AI Validation Requires Realistic Conditions

As AI infrastructure scales, workload and network emulation become central to validation. Few teams can recreate every production-scale topology, vendor combination, or workload scenario in a physical lab. Emulation helps model east-west flows, burst conditions, synchronization events, mixed traffic profiles, and congestion behavior before systems are deployed, while complementing existing compliance and interoperability testing.

Performance must also be characterized beyond peak throughput. Teams need to understand latency distributions, tail latency, congestion buildup, queue behavior, and performance degradation over time. In AI workloads, system efficiency is often dictated by the slowest GPU operations, making tail behavior a key metric.

Validation must also account for physical deployment realities. High-speed optics, dense cabling, thermal conditions, and power constraints can influence performance over time. Interconnect reliability therefore needs to be evaluated in ways that reflect handling, installation, sustained load, and environmental stress.

The urgency is especially clear in large-scale deployments, where issues need to be identified before infrastructure reaches production.

> *"[Because of] AI, [it] is becoming really critical to detect the issues before they touch production. If they go to production, it is already too late."*
>
> **Alessio Ferrari, Senior Cloud Optical Network Engineer, Microsoft**

The impact is both practical and economic. To keep pace with the demands of AI, validation must expose performance limits and interoperability risks earlier in the lifecycle. End-to-end validation helps reduce late-stage rework, minimize downtime risk, and protect the return on AI infrastructure investment.

What This Means for AI Infrastructure Readiness

Validation is no longer limited to proving that a single device, interface, or link works as expected. Risk now accumulates across the stack. A design margin issue at the physical layer, an interconnect sensitivity, a fabric congestion pattern, or a workload-driven timing issue may appear manageable on its own. When these conditions combine in a fully populated AI cluster, small weaknesses can become deployment-level constraints.

That is why readiness must be proven progressively, from physical interfaces to AI clusters. Each stage answers a different validation question and requires methods that reflect how that part of the infrastructure will behave before deployment.

- **Physical Layer:** Validate reliable operation at high bandwidth with real-time and sampling oscilloscopes with automated compliance workflows
- **Interconnects:** Evaluate performance under realistic conditions with Layer 2/3 network emulation
- **Network fabric:** Characterize throughput, latency, congestion, queue behavior, and mixed traffic across switches and racks
- **AI cluster:** Emulate workload traffic and scale-out behavior to identify bottlenecks and other performance issues before deployment

No single validation step can prove AI readiness by itself. Confidence comes from connecting results across the stack so teams can identify risk earlier, understand system performance issues, and evaluate infrastructure behavior before deployment.

Conclusion

What emerges from these industry perspectives is a clear message: AI infrastructure readiness requires validation across the full stack. Higher speeds, tighter margins, realistic AI traffic, and tightly coupled infrastructure are changing what teams must prove before deployment.

The path forward is a connected validation strategy that follows the infrastructure from physical-layer performance to interconnect integrity, fabric behavior, workload emulation, and AI cluster readiness. By recreating live production conditions earlier, teams can expose risks and build confidence before infrastructure is deployed.

Readiness is no longer defined by whether an individual component can pass compliance testing, but by whether the complete system delivers predictable, consistent performance under real-world conditions at scale.

Keysight enables innovators to push the boundaries of engineering by quickly solving design, emulation, and test challenges to create the best product experiences. Start your innovation journey at www.keysight.com.

KEYSIGHT

This information is subject to change without notice. © Keysight Technologies, 2026, Published in USA, October 14, 2026, 7126-1077.EN