Invisible Highways: Observability in the Evolution of AI-Era Networking Infrastructure

How network observability powers the backbone of artificial intelligence systems and why it matters for the future of high-performance computing.

Bhaskararao Vakamullu
Anna University
chennai India

From ARPANET to AI Enabler

1960s ARPANET

Pioneering packet-switching networks established the foundation of modern internet with rudimentary monitoring tools.

1980-2000 Enterprise

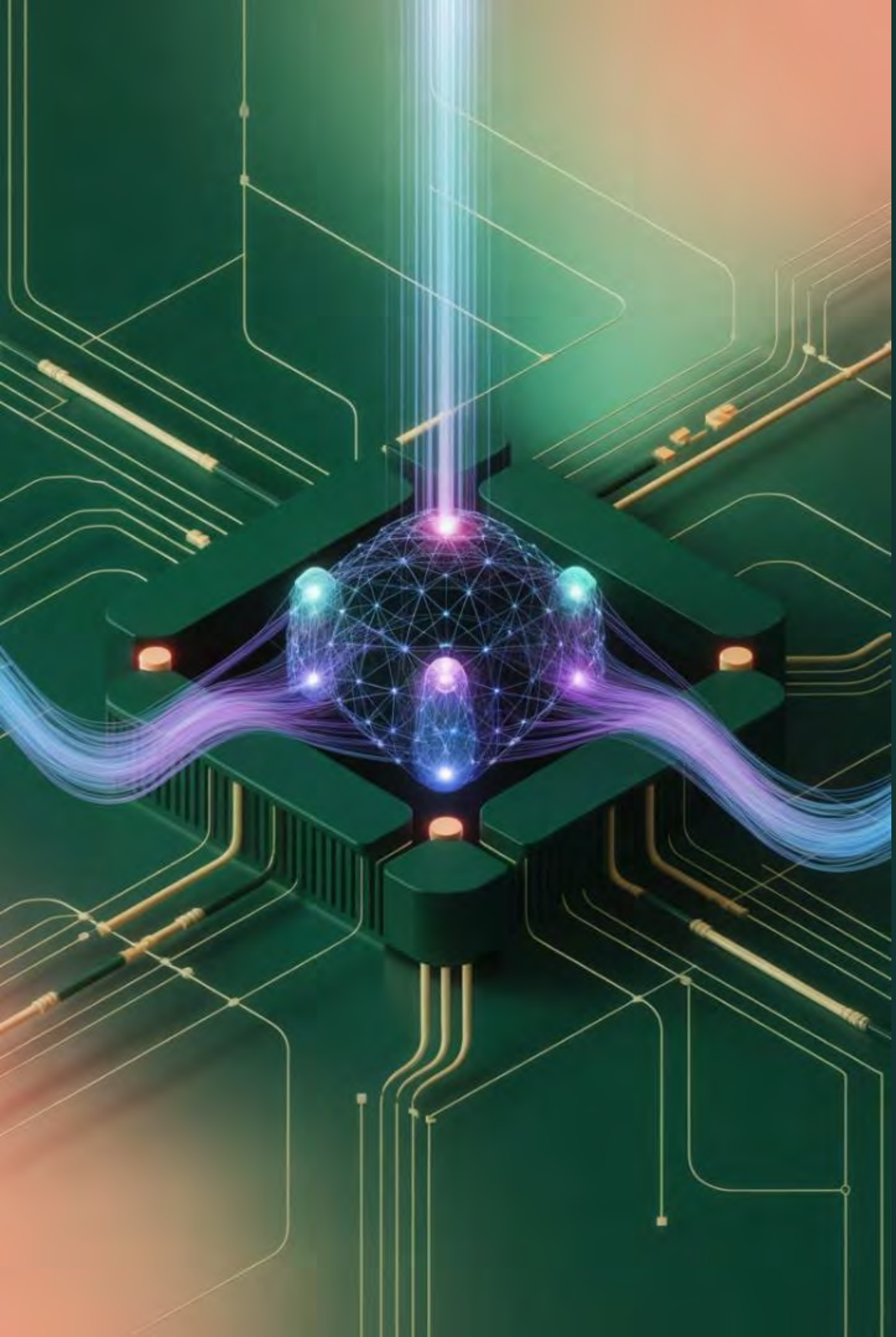
Evolution of Simple Network Management Protocol enabled standardized monitoring of network performance and health metrics.

2000-2015 Cloud

Software-defined networking revolutionized infrastructure with programmable control planes, enhancing network flexibility and visibility.

2015+ AI Era

Ultra-high bandwidth networks with sophisticated real-time telemetry systems now power distributed AI workloads at unprecedented scale.



Why Network Observability Matters for AI

Performance Guarantees

AI workloads demand consistent, ultra-low-latency communication across distributed compute clusters containing hundreds of interconnected nodes.

Failure Domain Isolation

Rapid identification and containment of network anomalies prevents costly cascading failures during critical AI training operations.

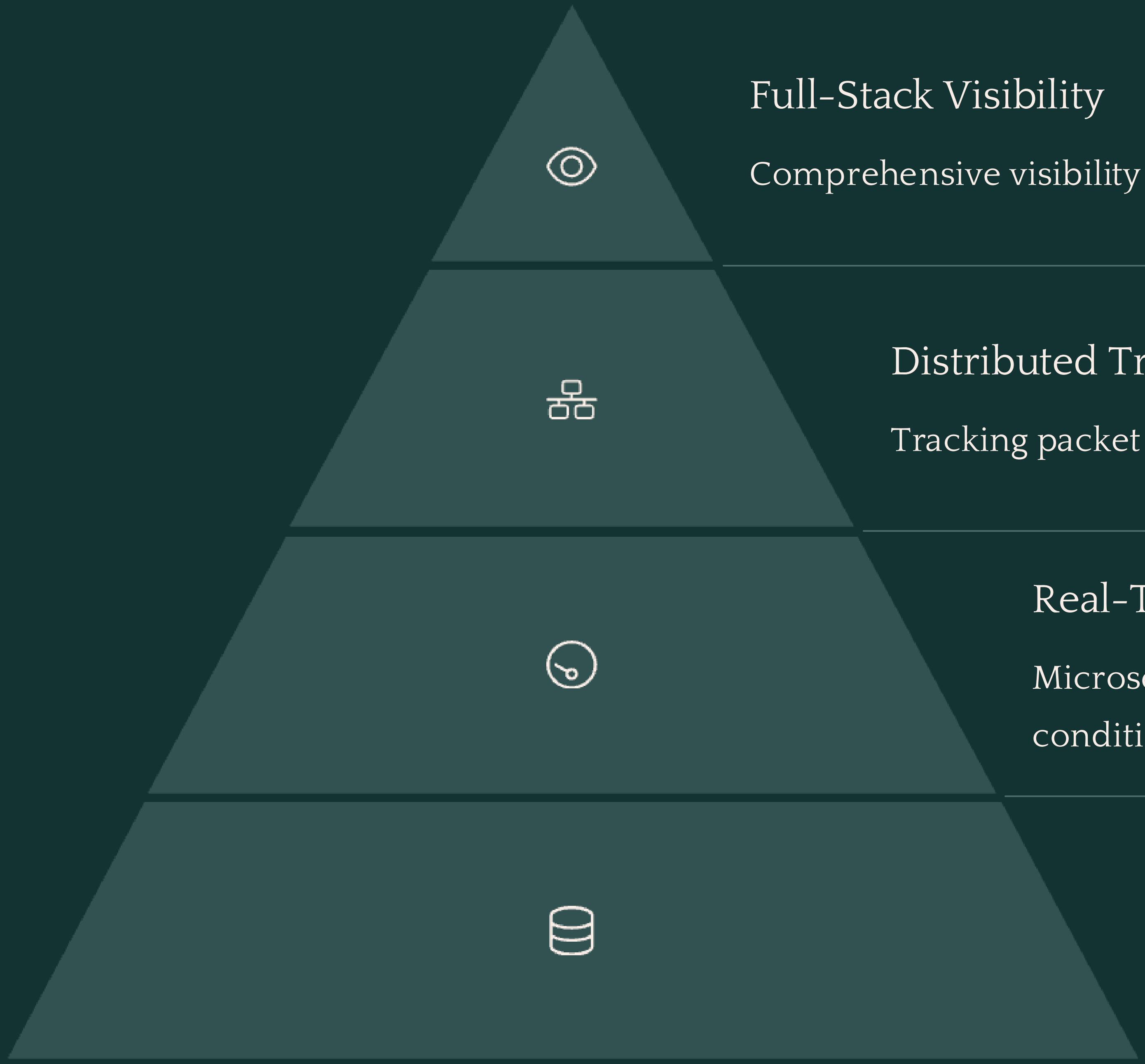
Resource Optimization

Deep visibility into traffic patterns and bandwidth utilization enables precise allocation of network resources for maximum AI computational efficiency.

Security Assurance

Sophisticated anomaly detection systems safeguard proprietary AI models and sensitive training datasets from exfiltration and tampering attempts.

The Invisible Scale Challenge



Full-Stack Visibility

Comprehensive visibility from application logic to physical infrastructure layers

Distributed Tracing

Tracking packet journeys across complex, interconnected AI computing clusters

Real-Time Telemetry

Microsecond-precision monitoring of critical network performance conditions

Hardware-Level Metrics

Detailed visibility into network ASIC buffer utilization and queue congestion patterns

Inside the Modern AI Network

400G+

Port Speed

Ultra-high bandwidth connectivity enabling massive parallel computing across GPU clusters

<1µs

Latency

Near-instantaneous response times critical for synchronized neural network training

32B+

Packets

Extraordinary volume of data packets processed daily within large-scale AI infrastructure

99.999%

Reliability

Five-nines availability essential for uninterrupted model training and inference operations

The Observability Triad

Metrics

- Real-time throughput monitoring and bandwidth utilization analytics
- Hardware-level queue depths and buffer allocation statistics
- Comprehensive error rate tracking and packet drop analysis
- Advanced link quality and signal integrity indicators

Logs

- Detailed control plane event recording and analysis
- Complete protocol negotiation and handshake audit trails
- Critical security incidents and access authorization events
- Automated configuration change tracking and validation

Traces

- End-to-end packet path visualization and routing analytics
- Distributed cross-node communication flow mapping
- Precise inter-hop timing and latency measurements
- Seamless correlation between network events and application activities

Programmable Telemetry Revolution

Traditional Monitoring

- Pull-based polling mechanisms (SNMP protocol)
- Coarse 5-minute collection intervals limiting responsiveness
- Fixed, predetermined counter sets with limited extensibility
- Significant CPU overhead impacting device performance
- Isolated data points with minimal cross-system correlation

Modern Telemetry

- Push-based streaming architecture (gRPC protocol)
- High-precision sub-second data resolution for real-time analysis
- Programmable, customizable data collection pipelines
- Efficient hardware-offloaded monitoring with minimal performance impact
- Rich contextual correlation enabling holistic system visibility

Self-Healing Network Architectures

Detect

Continuous high-resolution telemetry pinpoints performance degradations and anomalous patterns before they cascade into service disruptions.

Analyze

Machine learning algorithms correlate disparate events across the network fabric to isolate root causes with precision.

Execute

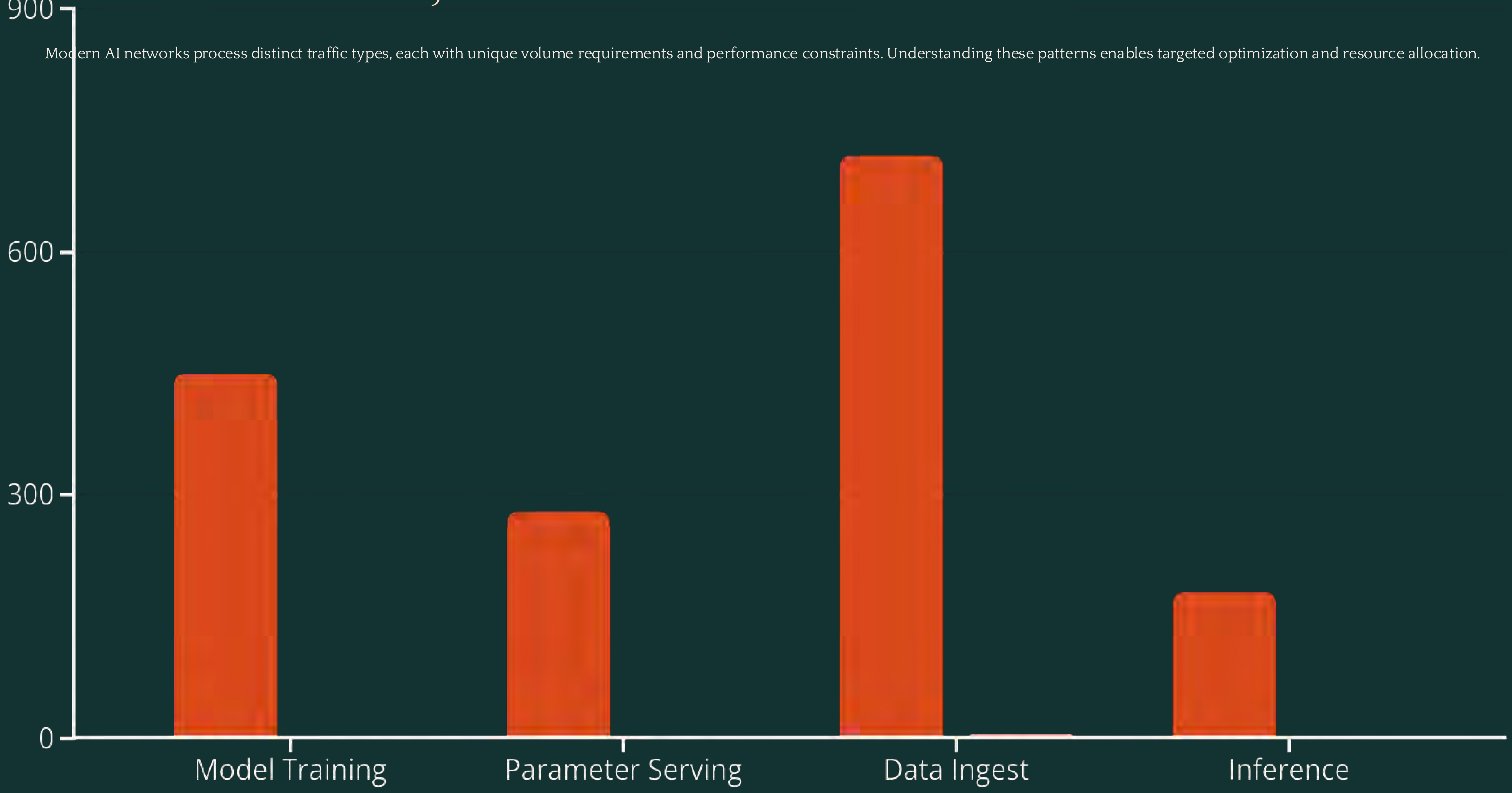
Orchestrated remediation workflows automatically implement corrective actions, verify their effectiveness, and restore network integrity without human intervention.

Decide

Intent-based policy frameworks evaluate multiple resolution pathways against business priorities to select optimal remediation strategies.

Fine-Grained Flow Analysis

Modern AI networks process distinct traffic types, each with unique volume requirements and performance constraints. Understanding these patterns enables targeted optimization and resource allocation.



Remaining Visibility Challenges

Hardware Opacity

Network ASIC internals function as black boxes, severely limiting visibility into critical packet processing decisions.

- Undocumented proprietary buffering mechanisms
- Vendor-specific optimization heuristics with unpredictable behaviors

Cross-Domain Correlation

Establishing clear connections between application performance and underlying network events demands sophisticated instrumentation.

- Non-synchronized timestamping across system components
- Absence of standardized contextual metadata for event correlation

Scale Limitations

Implementing comprehensive high-fidelity monitoring at AI infrastructure scale produces overwhelming telemetry volumes.

- Unsustainable storage requirements for complete historical data
- Significant computational burden for meaningful real-time analysis

Future of Observable Networks

AI-Powered Observability

Networks that monitor themselves using embedded ML agents.

Silicon-Level Telemetry

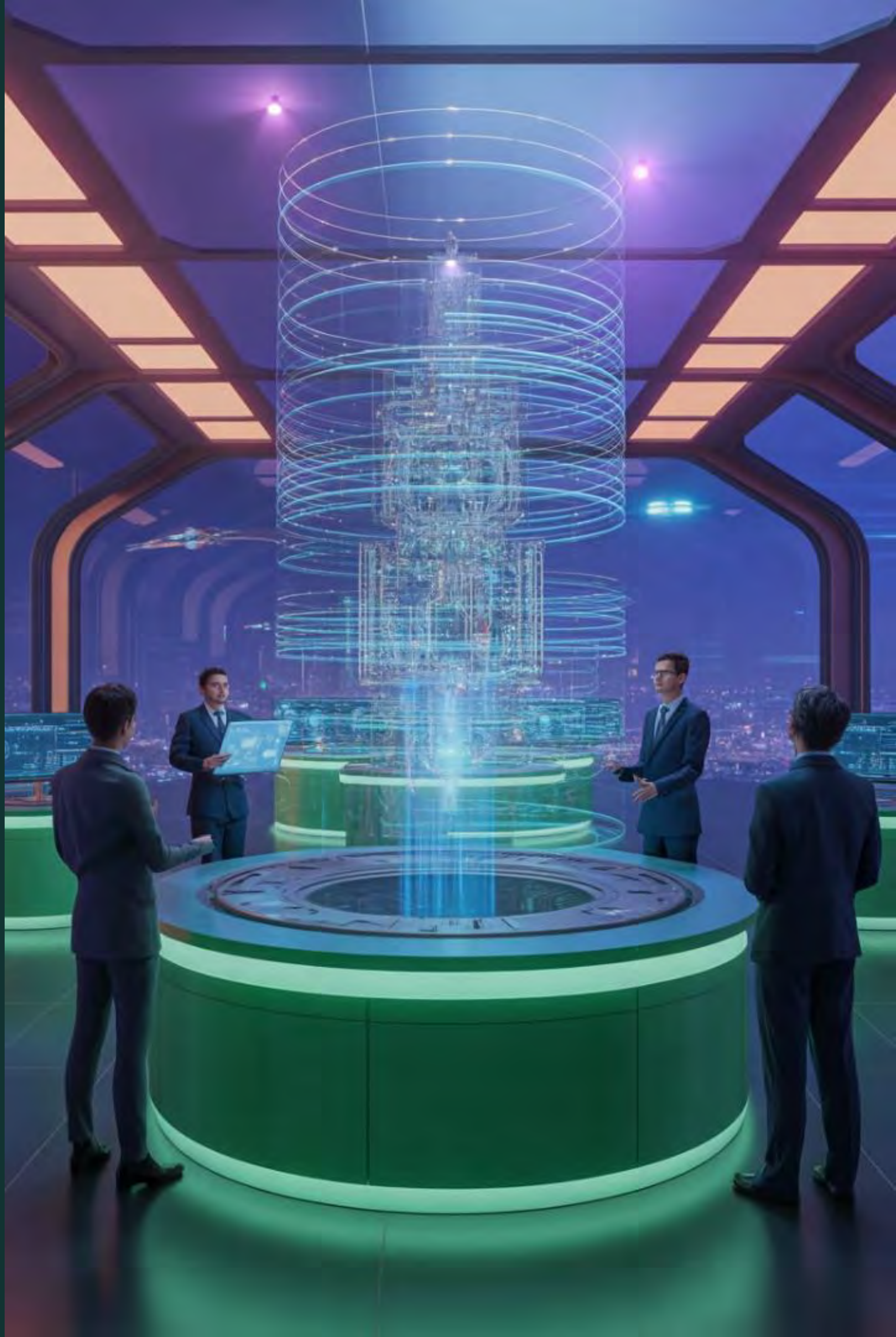
ASIC-integrated monitoring with zero performance impact.

Intent-Based Observability

Automatically translating business goals into monitoring policies.

Digital Twin Networks

Real-time simulation for predictive anomaly detection.



Thank you