AvidThink

AI in Networking 2025: Harnessing the AI Deluge

Transforming Networking Pipe Dreams to Reality with Agentic AI

RESEARCH BRIEF

Table of Contents

©2025 AvidThink LLC. All Rights Reserved.

RESEARCH BRIEF

AI in Networking 2025: Harnessing the AI Deluge

Transforming Networking Pipe Dreams to Reality with Agentic AI

Executive Summary

The AIOps landscape crossed a critical threshold in 2025. What began in 2024 as experimental generative AI chatbots powered by large language models (LLMs) has evolved into agentic AI frameworks capable of autonomous network operations in specific tasks. The breakthrough came from new reasoning models, which demonstrated human-like step-by-step problem-solving capabilities. These new models, combined with network control APIs, interoperability protocols (such as Anthropic's MCP and Google's A2A), and comprehensive telemetry, enabled vendors to build agentic systems that can autonomously diagnose network issues, generate remediation plans, and execute configuration changes with minimal human oversight — marking the industry's transition from basic automation (TM Forum Level 2) toward high autonomy (Level 4) in particular network domains.

Leading networking vendors have responded with a deluge of agentic frameworks, promising and demonstrating 30-50% savings in operational costs in early POCs. However, significant barriers remain. Trust in autonomous systems is a primary concern, as operators must navigate the risk that AI agents may make incorrect decisions, which could potentially lead to network outages. The potential for "agent sprawl" — hundreds of siloed, custom-built agents without unified governance — threatens to create new forms of technical debt. The human impact is equally profound, as network engineers evolve from hands-on operators to AI orchestrators who define business intent and manage fleets of intelligent agents, requiring new skills in programming, AI fundamentals, and data analytics. Meanwhile, the industry faces a "loss of expertise paradox" as veteran engineers retire.

The vision of autonomous networks now follows a more precise trajectory toward reality, emerging as a collection of autonomous domains rather than a single monolithic transformation. Organizations must shift their evaluation criteria from individual AI features to comprehensive agentic architectures, embracing "human-on-the-loop" deployments that build trust while enabling gradual transition to autonomous operations. The pipe dream of autonomous networking is rapidly becoming an operational reality — the industry's challenge is no longer whether AI can deliver autonomous networks, but how quickly organizations can adapt their people, processes, and architectures to harness this transformative capability safely and effectively.

Research Briefs are independent content created by analysts working for AvidThink LLC. These reports are made possible through the sponsorship of our commercial supporters. Sponsors do not have any editorial control over the report content, and the views represented herein are solely those of AvidThink LLC. For more information about report sponsorship, please reach out to us at research@avidthink.com.

About AvidThink

AvidThink is a research and analysis firm focused on providing cutting-edge insights into the latest in infrastructure technologies. Formerly SDxCentral's research group, AvidThink launched as an independent company in October 2018. AvidThink's coverage includes 5G infrastructure, enterprise networks, private wireless, edge computing, SD-WAN, SASE, ZTNA, cloud and AI infrastructure, and infrastructure security. Our clients range from Fortune 500 enterprises and hyperscalers to tier-1 communications service providers, fast-growing unicorns, and innovative startups. AvidThink's research has been quoted by Forbes, the Wall Street Journal, Light Reading, Fierce Networks, Mobile World Live, and other major publications. Visit AvidThink at avidthink.com.

©2025 AvidThink LLC. All Rights Reserved.

Introduction: From a Trickle to a Deluge of Intelligence

Looking Back: The 2024 Landscape of Pipe Dreams and AI Realities

Our 2024 report, "Pipe Dreams and AI Realities," captured the networking industry at the crest of the initial GenAI wave. Enterprises and communication service providers (CSPs) alike sought ways to apply the capabilities of LLMs across multiple facets of their businesses, driven by interest from C-level and board-level executives. They were sparked by the public's and financial markets' fascination with GenAI tools, such as OpenAI's ChatGPT and Anthropic's Claude. For networking, this translated into a flurry of activity centered on improving the human-computer interface (HCI) and streamlining operational workflows. Early dominant use cases included LLM-powered conversational assistants and knowledge retrieval systems for troubleshooting, which were primarily built on Retrieval-Augmented Generation (RAG) architectures.

These early implementations framed AI's role as an assistant or "copilot." The goal was to make network operators more efficient by allowing them to query network status in natural language, summarize complex alerts, or quickly find relevant information in vast documentation libraries. Predictive AI and GenAI were seen as complementary tools: the former for its proven value in forecasting, optimization, and anomaly detection, and the latter for its novel ability to understand and generate human-like text. While the long-term vision of fully autonomous networks remained the goal, the immediate, practical application of AI in 2024 focused on augmenting, rather than automating, human expertise.

The 2025 Reality: Rise of Reasoning and Agentic Workflows

Over the past 12 months, the AI technology landscape has matured rapidly. Just when LLM/FM (foundation model) performance began to plateau on human tasks, chain-of-thought prompting techniques, aided by inference-time scaling, ushered in a wave of reasoning models. Further application of post-training methods yielded OpenAI's ChatGPT o-1, o-3, Claude 3.5, and subsequently the DeepSeek R1 model, which demonstrated a remarkably human-like ability to reason through problems step-by-step[1].

To understand what these agents are capable of, we only have to look at the developer ecosystem and, to a lesser extent, what today's LLMs can do in terms of performing "deep research" — autonomously searching the web, collating information, and generating in-depth reports in answer to user queries. In software development, the rise of "vibe coding" and increasing capabilities of coding agents like OpenAI Codex, Claude Code, Gemini Code Assist, and agentic workflows built into IDEs like Cursor and Windsurf, as well as no-code coding platforms like Lovable, Replit, and Vercel v0, show the art of the possible. These developer-focused models have extensive datasets of existing code to learn from and, more importantly, automated methods to verify the correctness and performance of the resultant output. Such coding agents are now contributing to a significant proportion of new code created (30-50%) at major technology companies[2]. This type of exponential growth in agentic capability is captured in recent research from a non-profit organization, METR, which studies AI capabilities, particularly autonomous performance in R&D. One of their papers examines the performance of autonomous systems in terms of the length of tasks that AI agents can complete. The metric has been increasing exponentially, with a doubling time of around 7 months[3]. Unfortunately, network management and operations teams at enterprises and carriers neither have the same quality or richness of source data (which is usually not easily accessible or machine-consumable) nor an easy way to always verify the correctness of actions, such as for coding tasks. However, we can still look to the coding agent landscape as a sign of what is to come in agentic AIOps.

In late 2025, the conversation has already evolved from "What can GenAI do?" for network operations to "How can a multi-layered, multi-faceted AI architecture deliver measurable ROI for network operators?" Our 2025 reality involves an application of **hybrid intelligence** where **predictive AI**, **generative AI**, and a new, powerful layer of **agentic AI** work in concert. This rapid maturation from GenAI assistants to agentic AI frameworks demonstrates that the underlying technologies — robust APIs for network control, comprehensive streaming telemetry for observability, and powerful LLMs for reasoning — already laid the necessary groundwork. The primary barrier was not technological, but rather the conceptual framework for its application.

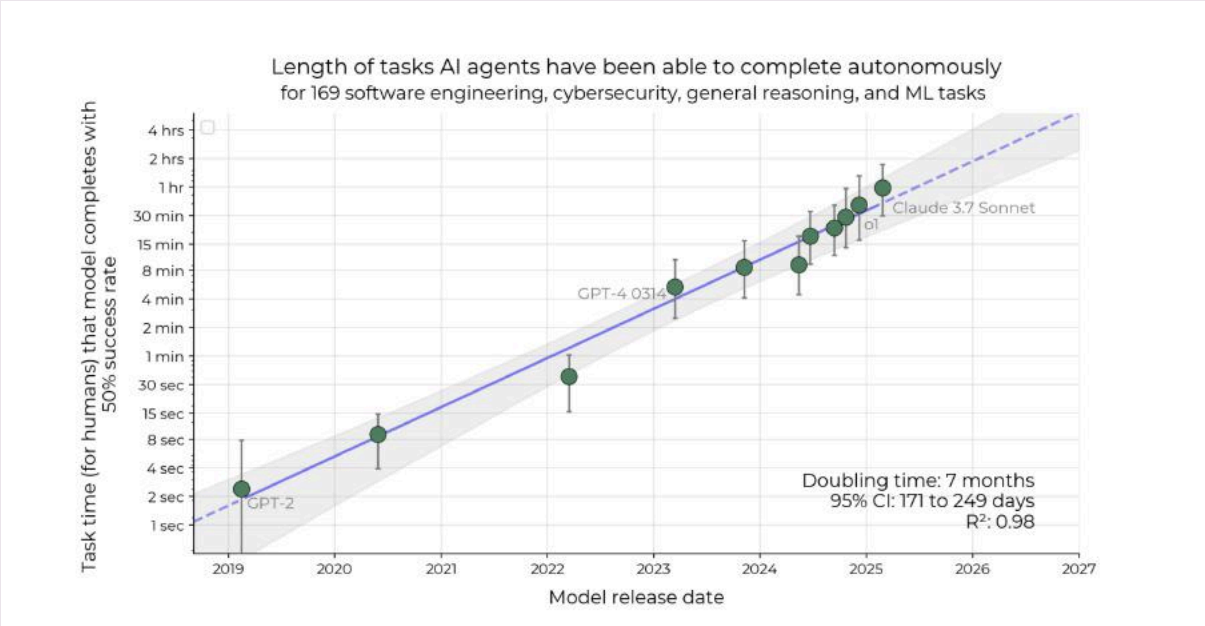
[1] There's active debate whether the models actually perform reasoning, or simply simulate reasoning steps. Regardless, the outcomes that these new models have achieved are remarkable and for this paper, we will use the term "reasoning" without strictly defining it.

[2] C. Mauran, "Mark Zuckerberg wants AI to do half of Meta's coding by 2026," Mashable, Apr. 30, 2025. https://mashable.com/article/llamacon-mark-zuckerberg-ai-writes-meta-code

[3] T. Kwa et al., "Measuring AI Ability to Complete Long Tasks," arXiv.org, 2025. https://arxiv.org/abs/2503.14499

©2025 AvidThink LLC. All Rights Reserved.

AI AGENT PERFORMANCE OVER TIME



AvidThink

Source: METR (Model Evaluation & Threat Research)

avidthink.com

Our 2024 report framed GenAI's role as an assistant, a passive tool that awaits and responds to human queries. In late 2025, leading networking vendors such as Amdocs, Cisco, Ciena Blue Planet, Ericsson, Extreme Networks, HPE Juniper, Huawei, Netcracker, Nokia, Ribbon Communications, and others flooded the networking market with their visions for "AgenticOps" and "Agentic AI Frameworks." This agentic breakthrough stems from reasoning models that can maintain focus on longer chains of reasoning. Utilizing these post-trained reasoning models, networking vendors have built AI systems that go beyond providing advice and can play the role of an autonomous actor within a defined workflow, capable of executing tasks and reacting to dynamic conditions, with humans acting as overseers.

The Autonomous Networking Journey: An Updated Roadmap

Revisiting the TM Forum Levels of Autonomy

The TM Forum's Autonomous Networks Maturity Model (see diagram on next page) continues to serve as the CSP standard guidepost for measuring the evolution from manual operations (Level 0) to a fully autonomous, self-governing state (Level 5). Even if enterprises don't always recognize or utilize the TM Forum framework, it's still a valuable and easy-to-understand tool for communicating autonomy levels. For example, CxOs from global CSPs stand on conference stages committing to their goals for autonomy — e.g., L4 by 2028 — and we're seeing finer-grained (even if somewhat ambiguous) goals in 2025, e.g., level 3.7, which highlights the framework's appeal.

The TM Forum framework, inspired by the SAE Levels of Driving Automation, provides a domain-agnostic language to describe the journey from basic, repetitive task automation (Level 1) to systems that can sense, analyze, and make decisions in complex, cross-domain environments (Level 4). And it serves as a convenient framework for our analysis of the industry's progress.

©2025 AvidThink LLC. All Rights Reserved.

TM FORUM LEVELS OF AUTONOMOUS NETWORKS

Autonomous Levels	L0: Manual Operation & Maintenance	L1: Assisted Operation & Maintenance	L2: Partial Autonomous Networks	L3: Conditional Autonomous Networks	L4: High Autonomous Networks	L5: Full Autonomous Networks
Execution	P	P/S	S	S	S	S
Awareness	P	P/S	P/S	S	S	S
Analysis	P	P	P/S	P/S	S	S
Decision	P	P	P	P/S	S	S
Intent/ Experience	P	P	P	P	P/S	S
Applicability	N/A	Select scenarios				All scenarios

P People (manual) S System (autonomous)

Source: TM Forum

avidthink.com

2025 Progress Report: Agentic AI as the Accelerator to L4/L5

Before 2024, CSPs and enterprises alike didn't have clear ideas on how to achieve L4 and beyond. However, the introduction and rapid maturation of agentic systems in 2025 have created a non-linear "jump" in autonomous capability, potentially making higher levels of autonomy achievable. Agentic AI addresses the core requirements of L4 (High Autonomous Networks), as defined by the TM Forum, which is a system capable of analysis and decision-making in a complex, cross-domain environment based on predictive or active, closed-loop management.

To achieve a specific goal, an AI agent needs to be able to (1) analyze a series of inputs and existing state, and (2) dynamically and autonomously figure out what it needs to do (which may involve using tools and resources to obtain more information), and then (3) take action. This aligns well with the TM Forum L4 model. Real-world examples emerging in 2025 demonstrate this in action. For example, the HPE-Juniper Networking portfolio features a Marvis Actions dashboard that supports the autonomous remediation of network issues such as misconfigured ports, capacity problems, and non-compliant hardware. This represents a step beyond L3 (Conditional Autonomy), where a system may optimize itself within a single domain under specific conditions, to L4 capabilities, where it actively analyzes and remediates complex issues.

This progress, however, is not yet occurring across entire network infrastructures (required to achieve L5). Instead, the industry can currently achieve high levels of autonomy only within specific, well-defined domains. A service provider might achieve L4 in their optical transport network using a vendor-specific solution from companies like Nokia or Ciena, while their enterprise campus network remains at L2 (or L3). This domain-specific advancement is a logical consequence of vendors focusing their deep expertise on particular segments of the networking ecosystem — Ericsson on the Wireless WAN, separate from its efforts on RAN, likewise with Nokia (data center separate from RAN), and Arista on data center first, followed by campus, before spreading to the branch via their recently-acquired VeloCloud assets.

The next major challenge for the industry is to move beyond autonomy within silos and orchestrate workflows across these domains, which may have varying levels of autonomy. This will drive cross-domain orchestration of agents, necessitating a control plane capable of managing end-to-end services across heterogeneous, autonomously managed domains. This is the strategic position that many enterprise and CSP vendors are marching towards in 2025 and beyond.

©2025 AvidThink LLC. All Rights Reserved.

The 2025 AI Toolkit for Networking: A Multi-Layered Architecture

Many of the topics we covered in our 2024 report remain pertinent in 2025. In particular, our section on the tasks and operations that can be performed by AI (predictive and generative) remains accurate and up-to-date. We refer the reader to the 2024 edition of this report for coverage of those topics. This year, we'll examine the evolution of major trends we saw in 2025, driven by the different flavors of AI, particularly the newer agentic forms.

Before we provide an update on the three different flavors of AI used in network operations, we will revise our AI Use Case Map (below) to account for the availability of agentic technology and modify the production use timeframes based on what we are seeing across enterprises and carriers. Many of the agentic use cases currently in development are still in limited production or are only used in proofs of concept (POCs), which we outline below.

AI USE CASE MAP CORE CAPABILITIES - UPDATED

	Discovery	Generation	Automation	Prediction & Simulation
Provides	Personalized Contextual Answers	Design, Configuration, Policy, Documentation	Tasks, Workflows	Analytics, Predictions, Optimization
Domains	Customer Service, Self Help	Day 0, 1, 2 Activities	Deployment, Customer Support, Troubleshooting	Network Optimization, Service Assurance, Planning, Security
Delivery	Intelligent Virtual Agent	Copilot	AI Agents	Control Modules, Copilot
Implementation	RAG	Fine-tuned FMs + Agentic	Fine-tuned FMs + Multi-Agent Agentic Workflow	Predictive, Generative and Agentic AI + Digital Twins
Effort	Low	Med-High	Med-High	Med-High
Timeframe (production-use)	Today	Today	6-12 months	Today and Evolving

AvidThink

avidthink.com

Predictive AI: The Maturing Foundation of AIOps

The foundation of any intelligent network operation remains the robust layer of predictive AI and machine learning (ML). As highlighted in 2024, predictive AI demonstrated its value in core AIOps tasks, such as anomaly detection, time-series forecasting, and pattern recognition. In 2025, this layer continues to hold value in these tasks and can act as input to the agentic systems operating above it. For example, systems that leverage ML to forecast network capacity requirements and proactively identify Quality of Experience (QoE) issues can utilize agentic AI to take action to mitigate these QoE problems before users experience degraded performance.

We're also seeing advances in the ML models used in time-series data, expanding from the typical recurrent neural networks (RNN) with long short-term memory (LSTM), or even convolutional neural networks (CNN), to variations like IBM Time Series Pulse (TSPulse)[4], and SoftBank Research's Unified Transformer Model[5] used in RAN forecasting and optimization. There's more cross-pollination between the latest advancements in GenAI and predictive AI, driving innovation and achieving new levels of performance.

[4] "IBM Granite Time Series Documentation – IBM Granite," Ibm.com, 2025. https://www.ibm.com/granite/docs/models/time-series/

[5] "SoftBank Corp. Boosts 5G AI-RAN Throughput by 30% with New Transformer AI | About Us | SoftBank," SoftBank, 2025. https://www.softbank.jp/en/corp/news/press/sbkk/2025/20250821_02/

©2025 AvidThink LLC. All Rights Reserved.

Generative AI: Evolving Beyond RAG

Retrieval-Augmented Generation (RAG) remains a cornerstone technique for grounding LLMs in proprietary, domain-specific data, but we're also seeing a move in 2025 towards diverse approaches in GenAI. The initial wave of RAG-powered chatbots is giving way to more advanced implementations and new use cases.

Advanced RAG techniques are emerging to address early limitations. Adding self-reflection to improve the RAG output (e.g., Self-RAG framework[6]), or leveraging knowledge graphs (e.g., Microsoft GraphRAG) to understand relationships between data entities, are means to improve retrieval accuracy and contextual understanding, providing more relevant output.

Beyond Q&A chatbots, GenAI is now being applied directly to network control and configuration. Vendors are demonstrating capabilities where GenAI can generate complete network device configurations from natural language-based intent, create complex automation scripts (such as Ansible playbooks or Python code), and even translate configurations between different vendors' command-line interfaces (CLIs). As we predicted last year for the 2025-2026 timeframe, this marks the beginning of GenAI's active participation in Day 0 and Day 1 operations.

The most significant trend under GenAI is the rise of domain-specific and task-specific models.

The most significant trend under GenAI is the rise of domain-specific and task-specific models. Recognizing that generic, large-scale FMs lack the nuanced understanding required for complex network operations, leading vendors are investing in training or fine-tuning their own foundation models on decades of proprietary data. Ericsson's research collaborations[7] shows the feasibility and potential resource savings of training smaller models (sometimes termed SLMs for "Small Language Models"). And Cisco's "Deep Network Model"[8] is a live service available to certified Cisco experts that's trained on an extensive corpus of technical assistance center (TAC) cases, knowledge base articles, and CCIE-level documentation to create a virtual expert with domain knowledge. Other vendors are taking smaller, open-weight models (e.g., Meta's Llama model family or Alibaba's Qwen) and performing post-training operations (such as supervised fine-tuning, distillation, etc.) to arrive at smaller, specialized models. This strategy can create a competitive moat, as the value of the AI becomes inextricably linked to the quality and uniqueness of the vendor's private dataset.

The New Frontier: Agentic AI and the Self-Driving Network

The most significant development in AIOps in 2025 is the emergence of agentic AI. Reasoning models enable the creation of autonomous agents that can consider context and state, and make informed decisions to achieve a desired outcome. Before reasoning models, it was unclear how we would achieve L4/L5 autonomy in and across multiple network domains, or accomplish the intent-based network automation that the industry has sought for over a decade.

To achieve agentic operation, many of the current agent-powered systems we see across enterprise and CSP vendors include three major components: memory (both short-term for context and long-term for learning), reasoning capabilities (the ability to break down a complex goal into a sequence of steps), access to tools and resources (via APIs or newer protocols like MCP and A2A) that allow it to interact with external systems and take action.

This has given rise to new architectural concepts, such as the "agentic AI mesh," where multiple specialized agents collaborate to solve complex, cross-domain problems. For example, Cisco's AgenticOps framework aims to have intelligent agents autonomously manage and optimize IT tasks across network, security, and observability domains. Similarly, HPE-Juniper is advancing its Mist platform to deliver agent-based AIOps through more autonomous and proactive network operations, shifting IT from a reactive to a proactive management approach. These frameworks enable a new class of use cases focused on proactive problem resolution, autonomous resource optimization, and dynamic security policy enforcement.

[6] A. Asai, Z. Wu, Y. Wang, A. Sil, and H. Hajishirzi, "Self-RAG: Learning to Retrieve, Generate, and Critique through Self-Reflection," arXiv.org, 2023. https://arxiv.org/abs/2310.11511

[7] Konstantinos Vandikas, "How much telecommunication knowledge can a small language model retain?," ericsson.com, Jun. 19, 2025. https://www.ericsson.com/en/blog/2025/6/how-much-telecommunication-knowledge-can-a-small-language-model-retain

[8] A. Raghavan, "Meet the Cisco Deep Network Model: Trained by the Experts, Purpose-Built for the Network," Cisco Blogs, Jun. 10, 2025. https://blogs.cisco.com/networking/meet-the-cisco-deep-network-model-trained-by-the-experts-purpose-built-for-the-network

©2025 AvidThink LLC. All Rights Reserved.

Digital Twins: From Simulation to Production ROI

Examining other key developments outside the trio of AI model categories, we observe that the role of digital twins in networking has evolved from a promising concept for risk-free simulation into a production-grade technology delivering real-world value. In particular, digital twins can be used by agentic systems to validate pending changes with minimal risk. While our 2024 report highlighted the potential of digital twins for modeling and "what-if" analysis, 2025 brings numerous concrete case studies that demonstrate tangible results. Examples include AWS with Orange[9] and NTT[10] for root cause analysis, as well as another use of digital twins for optimization by Chunghwa Telecom in Taiwan, in collaboration with Ericsson,[11] which resulted in a 14% increase in capacity and improved service reliability during a New Year's Eve celebration.

Other industry examples further solidify this trend. Nokia has reported using digital twins to reduce network deployment cycle times by 30%[12]. These successes demonstrate that digital twins are a crucial component of the modern AIOps toolkit, enabling operators to mitigate risks, optimize performance, and validate the actions of AI agents before they are executed on the live network.

MCP, A2A — Interoperability Protocols Powering Agentic AIOps

Reasoning models enable the creation of autonomous agents that can consider context and state, and make informed decisions to achieve a desired outcome.

The scaling of agentic AI operations in networking will depend on standardized protocols that enable AI agents to discover, authenticate, communicate, and collaborate across different vendors and frameworks. While several candidates have emerged, two prominent protocols are Anthropic's Model Context Protocol (MCP) and Google's Agent2Agent (A2A).

Model Context Protocol (MCP), developed by Anthropic, serves as the connectivity layer between AI models and external data sources, APIs, and tools. MCP enables AI agents to access real-time context, including network topology, device configurations, and operational state, ensuring that generative AI responses are grounded in network context rather than providing generic recommendations. This protocol can support configuration generation and troubleshooting workflows where accuracy and environment-specific awareness are essential. MCP has seen widespread adoption across SaaS applications and the enterprise ecosystem, and many networking vendors have embraced creating MCP Servers to enable more flexible and no-code API integration with their platforms. While not a silver bullet and somewhat overhyped, we envision MCP playing a critical role in the agentic landscape. One huge caveat for the use of MCP is the lack of security considerations in its initial incarnations. It was designed to be flexible and easy to use, and the community is working hard to add authentication, authorization, and auditability, as well as improved policies and controls. Many exploits and weaknesses are still inherent in MCP, but we expect the community to rapidly close the security gaps given its criticality to the agentic landscape outside networking.

Agent2Agent (A2A), contributed to by Google through the Linux Foundation in 2025, is another critical protocol to watch. A2A provides a common language that AI agents use to discover and delegate tasks to one another across different platforms and frameworks. A2A addresses the fundamental challenge of enabling agents to work together in dynamic, multi-agent environments, coordinating actions across a wide array of applications and data infrastructures. The protocol has garnered support from over 100 technology companies and serves as a critical building block for enterprise-scale agentic deployments.

---

[9] "Networks for AI and AI for Networks: AWS and Orange's Journey - Hello Future," Hello Future, May 02, 2025. https://hellofuture.orange.com/en/networks-for-ai-and-ai-for-networks-aws-and-oranges-journey/

[10] "Beyond Correlation: Finding Root-Causes using a network digital twin graph and agentic AI | Amazon Web Services," Amazon Web Services, Aug. 18, 2025. https://aws.amazon.com/blogs/database/beyond-correlation-finding-root-causes-using-a-network-digital-twin-graph-and-agentic-ai/

[11] "Chunghwa Telecom and Ericsson harness AI and digital twins to secure network performances during data surge," ericsson.com, Feb. 20, 2025. https://www.ericsson.com/en/press-releases/2/2025/2/chunghwa-telecom-and-ericsson-harness-ai-and-digital-twins-to-secure-network-performances-during-data-surge

[12] "Speed up your network deployments by 30% with digitalization | Nokia.com," Nokia.com, 2023. https://www.nokia.com/blog/speed-up-your-network-deployments-by-30-with-digitalization/

©2025 AvidThink LLC. All Rights Reserved.

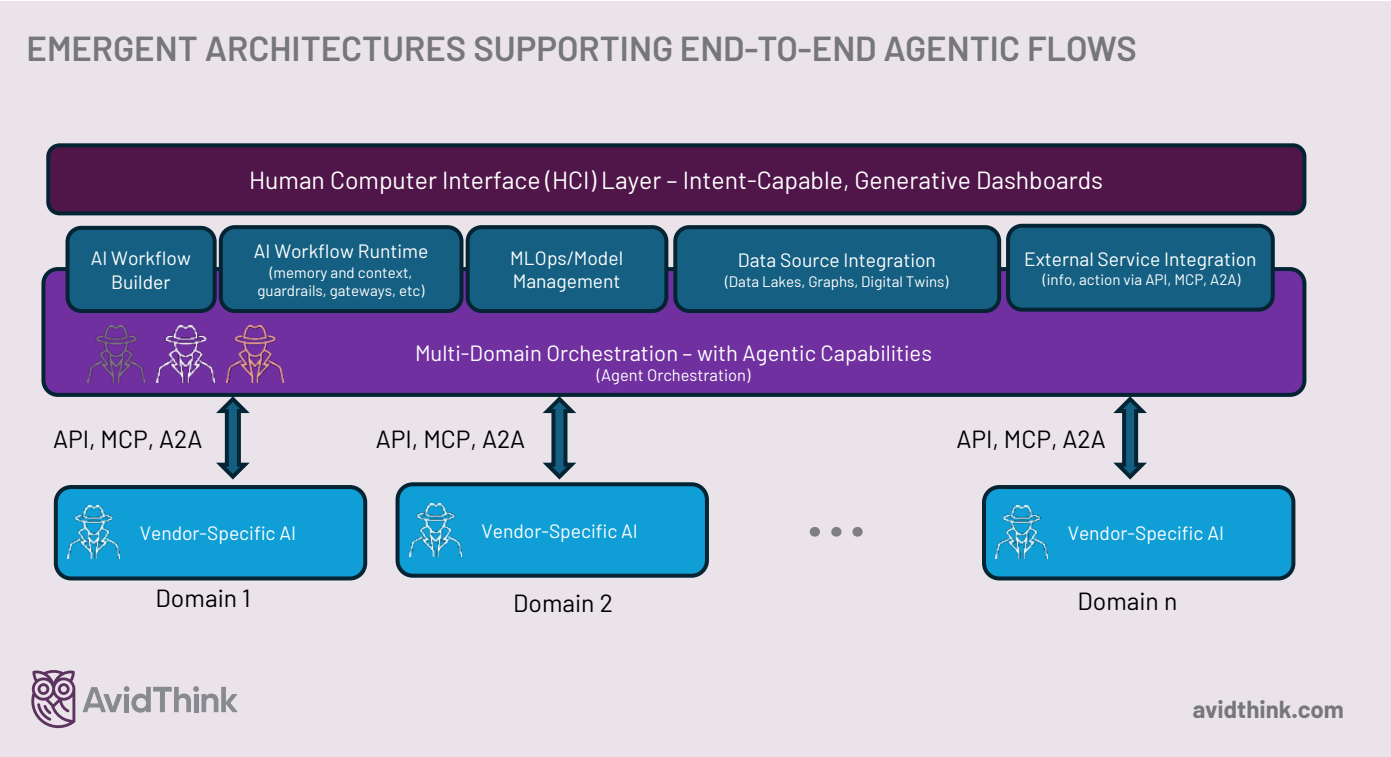
While other agentic protocol candidates exist, MCP and A2A have established themselves as the dominant protocols. For example, a prominent candidate, Agent Communication Protocol (ACP), developed through IBM's BeeAI project and hosted by the Linux Foundation, has now been merged with A2A[13].

As for agentic platforms and frameworks, AGNTCY, initially developed by Cisco's Outshift incubator and contributed to the Linux Foundation in July 2025[14], provides a comprehensive multi-agent infrastructure layer that encompasses agent discovery, identity, messaging, and observability. Its core capabilities include the Open Agent Schema Framework (OASF) for DNS-like agent discovery, cryptographically verifiable identities for secure cross-organizational operations, SLIM (Secure Low-latency Interactive Messaging) for quantum-safe communications, and specialized observability SDKs for monitoring complex multi-agent workflows.

In the long term, the success of these protocols and frameworks in helping establish autonomous networking operations will depend on their ability to maintain security, reliability, and performance at enterprise and carrier scale while continuing to foster an open, vendor-neutral ecosystem.

Putting it Together: Emergent Architectures and Practices

The incorporation of agentic AI is helping to crystallize a common architectural pattern that networking vendors are implementing with some variations across their platforms. As illustrated in the diagram below, this emergent architecture represents a multi-layered approach that combines domain-specific intelligence with cross-domain orchestration capabilities. At the foundation, vendor-specific AI systems operate within distinct network domains — whether campus, data center, WAN, transport, core, or RAN — leveraging hybrid intelligence that combines predictive AI for forecasting and optimization, generative AI for configuration and troubleshooting assistance, and agentic AI for autonomous decision-making. Notably, many of these systems we've seen retain some traditional rules-based control mechanisms where AI is not the optimal solution, recognizing that not every network operation benefits from machine learning approaches.



[13] "ACP Joins Forces with A2A – LFAI & Data," Lfaidata.foundation, 2025. https://lfaidata.foundation/communityblog/2025/08/29/acp-joins-forces-with-a2a-under-the-linux-foundations-lf-ai-data/

[14] "Outshift | AGNTCY project donated to Linux Foundation with major industry backing," Outshift by Cisco, 2022. https://outshift.cisco.com/blog/agntcy-donated-to-linux-foundation

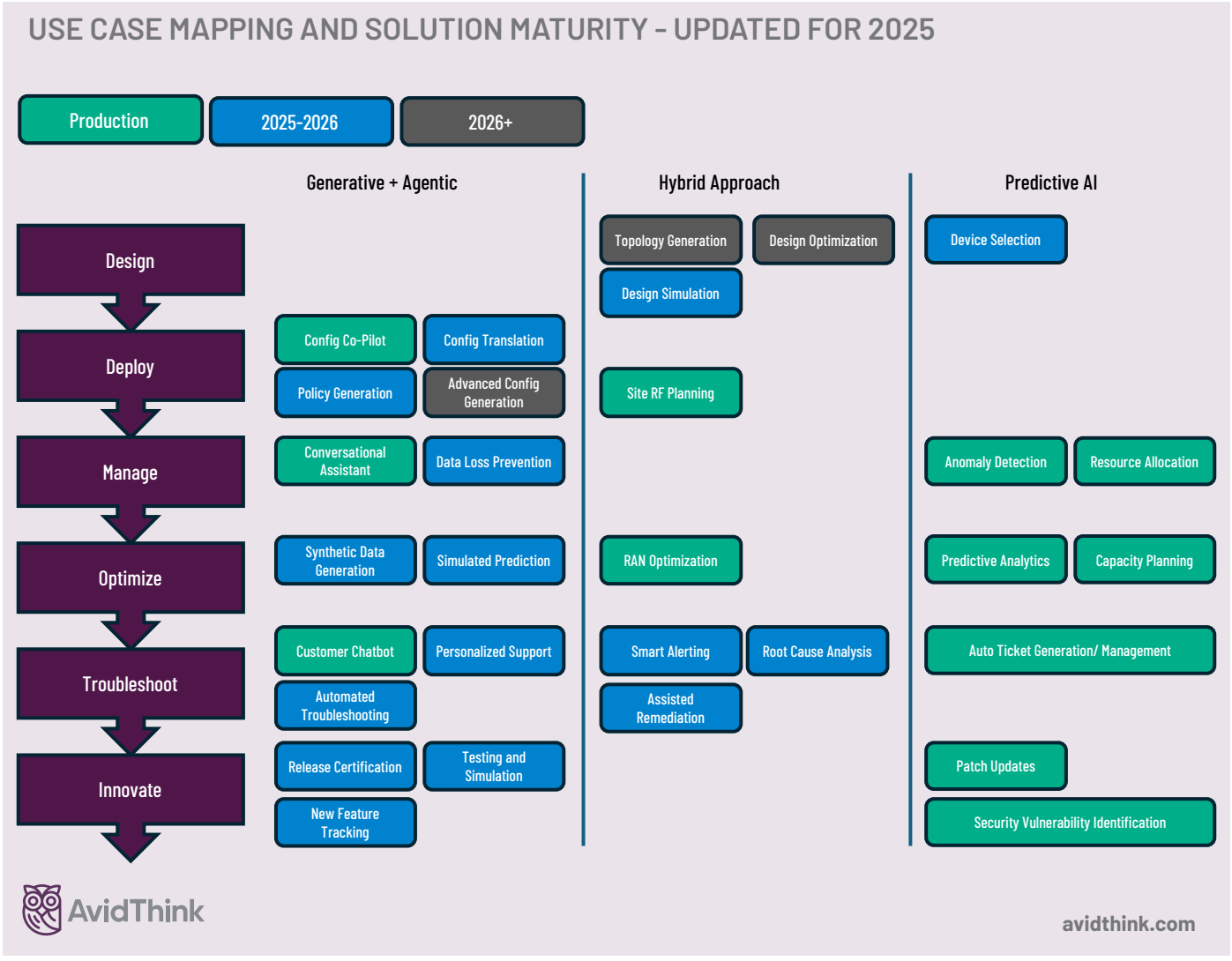
©2025 AvidThink LLC. All Rights Reserved.

The architecture, unsurprisingly, mirrors the evolution of Multi-Domain Service Orchestrators (MDSOs) that carriers have deployed, but now enhanced with agentic capabilities. This orchestration layer enables coordinating agents with higher-level service objectives to interact seamlessly with domain-specific agents optimized for particular tasks — such as route configuration, Wi-Fi optimization, or RAN parameter tuning. The standardization of interoperability protocols, such as MCP and A2A, will prove helpful in enabling agent-to-agent communication, allowing leading vendors like Cisco and HPE-Juniper to demonstrate end-to-end service troubleshooting that spans multiple network domains and vendor boundaries. What we expect to emerge is not a monolithic autonomous network, but a federation of specialized autonomous domains working in concert.

This architectural pattern addresses the practical reality that autonomous networking will not arrive as a single transformation but as a collection of progressively more intelligent domains. The unified data foundation, incorporating data lakes, digital twins, and real-time telemetry, provides the contextual awareness that enables both domain-specific and cross-domain agents to make informed decisions. Meanwhile, the human-computer interface layer is evolving beyond simple chatbots to intent-capable, generative dashboards that will allow network operators to define business objectives at a high level. At the same time, the underlying platform translates these intents into coordinated actions across the network infrastructure. This emergent architecture represents the industry's practical path toward the long-promised vision of self-operating networks.

AI's Evolving Impact Across the Network Lifecycle - Augmentation to Autonomy

The introduction of more advanced agentic capabilities in 2025 requires a re-evaluation of AI's role across the Day 0, 1, and 2+ operational lifecycle. The 2024 assessment, which focused on AI-assisted and copilot functions, is now augmented by a new class of autonomous actions. For example, Day 0 network design is no longer "AI-assisted"; it can now involve AI agents



©2025 AvidThink LLC. All Rights Reserved.

generating and simulating multiple complete topology options based on high-level business intent, optimizing for cost, performance, and resiliency. Similarly, Day 2 troubleshooting is evolving from a GenAI assistant providing suggestions to an AI agent that can autonomously diagnose the root cause, propose a remediation plan, and, upon approval, execute the necessary configuration changes.

The updated chart on the previous page provides our current outlook on the maturity of key AI-driven use cases, reflecting accelerated timelines and new capabilities that have emerged since the 2024 report. It serves as a high-level guide to the current state and future trajectory of AI in practical network operations.

Our updated mapping highlights several critical shifts. The most significant is the use of agentic capabilities across a wide range of Day 1 and 2 tasks, including root cause analysis and automated troubleshooting. While they appear similar on the map to last year's, the capabilities represent a different class of automation than the "assisted" functions envisioned for 2024. They are now available in early adopter releases from leading vendors. Timelines for previously anticipated features have also been accelerated. Conversational assistants, projected as a 12-month future item in 2024, are now a mainstream, table-stakes feature across the industry.

2025 Vendor Landscape: The Race to Agentic

The AI in the networking market in 2025 is defined by a race to build unified, AI-native, cross-domain network orchestration and management platforms. The following provides a current profile of the key companies shaping this dynamic landscape.

ARISTA

Arista is advancing its AI-driven networking capabilities through its Autonomous Virtual Assist (AVA) service, positioning itself as a comprehensive AIOps solution provider. Powered by Arista's EOS networking software and integrated into CloudVision, AVA draws intelligence from a multi-modal data lake, NetDL (Network Data Lake), that collects high-fidelity telemetry not only from Arista devices across data center, campus, and WAN domains, but also from third-party network elements, virtualization platforms like VMware, and compute infrastructure. This cross-domain approach enables Arista to provide 360-degree network observability under its CV UNO (CloudVision Universal Network Observability) license. AVA encompasses several key components: Ask AVA, a conversational chatbot interface for natural language network queries; AVA Insights for proactive issue detection and root cause analysis; and AVA Agents for faster, automated troubleshooting workflows.

Arista's recent acquisition of VeloCloud from Broadcom strengthens its AI portfolio, bringing VeloCloud's AIOps capabilities to SD-WAN environments, including AI-powered application identification for encrypted traffic and automated fault analysis. The combined Arista-VeloCloud offering aims to unify operations across branch office switching, wireless, and WAN connectivity, providing seamless cross-domain visibility and AI-driven insights from edge to data center. Additionally, Arista strengthens preventive operations through AVD (Arista Validated Design), a network automation framework built on Arista Continuous Integration (CI) Pipeline. When integrated with CV UNO, AVD can leverage its AVA-powered application awareness to predict potential impacts of network changes before deployment, minimizing the risk of production issues.

blueplanet
a division of Ciena

Many CSPs view **Blue Planet** as a specialized provider of AI-driven network automation solutions designed for telecommunications service providers modernizing their legacy OSS environments. At the core of its AI strategy is the Blue Planet Agentic AI Framework, a purpose-built system that operates across three layers: the agentic core for building, running, and orchestrating AI agents; agentic tooling that provides agents access to comprehensive OSS data and APIs spanning inventory, orchestration, and assurance; and agentic channels which provides application-level integration where agents deliver enhanced capabilities directly within Blue Planet's portfolio. Blue Planet AI Studio complements this framework. This integrated development environment enables CSPs to build, test, and deploy custom AI agents, supporting a "bring your own AI" model that prevents vendor lock-in and addresses data sovereignty concerns.

©2025 AvidThink LLC. All Rights Reserved.

Blue Planet's AI implementation focuses on both optimizing existing network processes and enabling transformational approaches to service delivery. Key use cases include 5G network slicing with natural language intent conversion, dynamic service catalog creation, and AI-powered network planning that combines conversational interfaces with graph neural network technology for predictive impact analysis. The company emphasizes that its Blue Planet Inventory (BPI) solution serves as the essential data foundation, creating a federated single source of truth across network layers and vendor environments. Their deep telecom domain expertise, built on years of cloud native OSS application development, allows Blue Planet's AI agents to leverage comprehensive network topologies, routing protocols, and orchestration capabilities.

Cisco is repositioning its network management and automation offerings as a suite of comprehensive agentic operations through its newly launched AgenticOps framework, introduced in June 2025 (at Cisco Live US). At the heart of this is the Cisco AI Assistant. This natural language interface enables administrators to diagnose network issues, automate complex workflows, and identify root causes through conversational interactions. This is complemented by Cisco AI Canvas, a generative UI workspace that unifies networking, security, and observability domains under a single intelligent interface. These capabilities rest on Cisco's Deep Network Model (covered earlier).

Meanwhile, Cisco's end-to-end assurance strategy has been enhanced through innovations to its ThousandEyes product line, announced at Cisco Live 2025, which brings AI-powered visibility deeper into enterprise environments and improves operational insight across complete digital experiences. The new bidirectional integration between Splunk and ThousandEyes creates a unified observability and assurance platform, where ThousandEyes extends network intelligence into Splunk for enhanced service health correlation. Meanwhile, Splunk Observability Cloud provides distributed tracing support, showing backend application services within the ThousandEyes platform.

Outside product capabilities, Cisco has invested deeply into agentic communications protocols A2A and MCP via their AGNTCY initiative (covered earlier). Cisco is also positioning itself for the anticipated surge in "agentic AI" traffic patterns, which differ from traditional network loads in that they are more persistent and continuous, rather than spike-driven.

Ericsson Enterprise Wireless Solutions (EEWS) leverages its NetCloud management platform to bring orchestration and AIOps capabilities to Wireless WAN and Private 5G enterprise solutions. EEWS recently announced new agentic capabilities for its AI virtual expert, ANA (AI-based NetCloud Assistant). ANA is designed as an agentic framework with multiple orchestrators and functional AI agents, capable of executing complex workflows and learning in real-time. Key AI features include automated troubleshooting workflows that will address the top issues identified by Ericsson support teams, partners, and customers. ANA also provides multi-modal content generation with dynamic graph creation and explainable AI capabilities that display real-time process feedback. The company plans a phased rollout starting with a troubleshooting orchestrator in Q4 2025, followed by configuration, deployment, and policy agents in 2026.

Beyond Enterprise Wireless Solutions, **Ericsson** has been advancing AI across their broader portfolio with major carrier deployments. Ericsson is leading the network management transformation with an extensive rApps (RAN Intelligent Controller non-real-time applications) portfolio that embeds AI/ML-powered, policy-driven intelligence directly in RAN. Major customers include NTT DOCOMO, which selected Ericsson and the AI-based Performance Diagnostic solution for optimization of its nationwide radio access network (RAN), and Swisscom, which relies on Ericsson AI solutions to manage network complexity. Ericsson has also broadened its AI-RAN ecosystem by partnering with startups like Aira Technologies, which leverage Ericsson's Intelligent Automation Platform (EIAP). AT&T, Aira Technologies, and Ericsson achieved a world-first by deploying a generative AI-generated rApp in AT&T's test environment in June 2025. Ericsson has also partnered with AWS to launch the Gen-AI Lab for communication service providers (CSPs) worldwide, with Grameenphone as the first CSP to create an Agentic-AI-led solution.

©2025 AvidThink LLC. All Rights Reserved.

HPE Aruba Networking emerged from the July 2025 completion of HPE's acquisition of **Juniper** with an accelerated roadmap toward building a "secure AI-native network" that leverages the combined strengths of both companies' AI portfolios. The enhanced Mist platform, originally from Juniper, now features agentic AI capabilities centered around an improved Marvis AI engine and Marvis AI Assistant. Innovations include the Marvis Conversational Interface, which enables natural language queries such as "Why is this site slow?" with intelligent explanations, and Marvis Minis (digital twins) that proactively simulate user experiences to predict issues before they manifest. The platform's purported crown jewel is the Marvis Large Experience Model (LEM), which combines telemetry data with advanced attribution models to quantify how various network elements — clients, access points, devices — contribute to degraded user experiences.

The unified entity is advancing beyond reactive troubleshooting toward autonomous remediation through bounded agentic actions, which include automatic fixes for VLAN misconfigurations, loop detection and port shutdown, non-compliant device upgrades, stuck port resolution, and enforcement of policy/firmware compliance. HPE's vision encompasses deep integration between its Aruba networking and Juniper product lines under a common cloud platform, AI engine, operating system, and hardware architecture, creating a unified AI-native umbrella. While the company benefits from Juniper/Mist's mature AI/ML foundation in wireless environments and extensive customer footprint in distributed deployments, challenges remain in extending generative and agentic features to more constrained domains such as core, backbone, and edge networks.

Huawei has positioned itself at the forefront of autonomous networking through ambitious claims around its AI Core Network, launched at MWC Barcelona 2025 as an "autonomous generative network capable of self-optimization and self-O&M (Operation & Maintenance)." The company's roadmap unfolds in two phases: Phase 1 introduces the "5G-A Intelligent Core," which features integrated AI agents. And Phase 2 delivers the "Agentic Core" leveraging Huawei's proprietary AI architecture (AIBA) to achieve self-optimizing and self-operating capabilities. This vision is supported by the upgraded Xinghe AI Fabric solution for enterprise networks, which incorporates network-scale load balancing (NSLB) to optimize path planning, eliminate performance bottlenecks, and enhance reliability through intelligent automation.

The foremost concern we've heard from enterprises and carriers is Trust and Control in Agentic Systems. The prospect of AI agents autonomously executing changes on a production network requires careful risk management.

Huawei's AI-centric approach extends across its 5G-Advanced (5G-A) strategy, emphasizing the automation of operations and maintenance, high-level network autonomy, and intent-driven automation capabilities, as demonstrated through initiatives like GreenPulse. The company is investing in telecom-specific foundation models and digital twins to enable the simulation of network behavior, predictive planning, and autonomous operation. Key implementations include RAN Intelligent Agents, role-based AI copilots, and scenario-based solution automation designed for 5G-A networks. However, Huawei may face challenges in integrating these advanced capabilities across legacy network infrastructures, ensuring reliability in mission-critical core, backbone, and edge interoperability scenarios, addressing security and privacy concerns inherent in generative and agentic AI deployments, and navigating competitive and regulatory pressures that limit market access in key regions.

IBM is positioning itself as a comprehensive AI platform provider for networking through its newly launched IBM Network Intelligence product. The company's approach centers on what it calls a "dual-intelligence" architecture that combines analytical AI for handling network telemetry at scale with generative AI for contextual reasoning and automation. This new effort by IBM complements its expansive AI watonsx portfolio of products which enable organizations to build trusted and reliable AI systems. IBM has also developed open-source time series foundation models, including IBM Time Series Pulse and IBM Tiny Time Mixers, which are compact, pre-trained models with 1-5 million parameters that work well across various network domains, including IP, RAN, SD-WAN, and data center. These models utilize a unique TS Mixer architecture that enables simultaneous multivariate anomaly detection across time and frequency domains. They have been productized for streaming data and achieved 1,000x inference performance improvements, in the order of milliseconds.

©2025 AvidThink LLC. All Rights Reserved.

IBM's Network Intelligence platform features a sophisticated agentic framework powered by reasoning LLMs that perform iterative root cause analysis using a ReAct pattern with chain-of-thought reasoning. The agents operate through multiple filtering layers, form hypotheses about root causes, and actively collect data on demand to prove or disprove their theories through an iterative loop of hypothesis generation and validation. The new system utilizes IBM's Concert Workflows technology. It integrates with Red Hat Ansible and HashiCorp Terraform, providing embedded automation capabilities that enable the dynamic invocation of thousands of APIs from hundreds of vendors through function-calling agents. IBM emphasizes explainability and trust through human-in-the-loop control and comprehensive scaffolding around the LLMs, including detailed reasoning traces and standardized prompts. This approach incorporates reinforcement learning, allowing the system to improve over time based on operator feedback and subject matter expert refinements to the agent's reasoning processes.

Nokia has positioned itself as a pioneer in autonomous networking through its "Sense, Think, and Act" vision, targeting TM Forum Level 4 autonomy and beyond. Central to Nokia's strategy is building trust in AI-driven operations through multiple validation mechanisms, including digital twins that serve as network emulation layers for testing candidate configurations. Nokia's latest agentic capabilities encompass network operations, featuring Digital Operations Center automation for service orchestration and troubleshooting, as well as MantaRay, Altiplano, WaveSuite, NSP, and EDA for various network domains. They also cover AI-powered security features, such as the recently introduced Threat Hunt Assistant from Netguard Cybersecurity Dome, which reduces threat dwell time, as well as enhanced subscriber experience analytics with natural language interfaces. Nokia Deepfield leverages AI-driven big data technology to detect and mitigate DDoS attacks.

The company's hybrid AI strategy recognizes that not all network operations benefit from AI — proven programmatic solutions; BGP control, for example, remains untouched. At the same time, AI excels in areas like troubleshooting, anomaly detection, and predictive maintenance. Nokia's AI capabilities span across its entire portfolio, from AI-powered RAN optimization that improves 5G mobile radio network performance and energy efficiency, to fixed broadband network automation, and IP, optical, and data center networks for cloud and AI infrastructure. Through partnerships with hyperscalers like AWS and Google Cloud, Nokia is accelerating the deployment of telco-trained models, digital twins, and AI agents that help operators detect network anomalies faster, anticipate service-affecting issues, and optimize resource utilization across all network domains.

Ribbon Communications has carved out a strategic position by focusing on AI-driven analytics, security, and assurance for real-time communications (RTC) networks. Their target markets include service providers, large enterprises, and critical infrastructure and government sectors, such as the US Department of Defense (DoD). Ribbon recently announced Acumen, its new AIOps & Automation platform, which distinguishes itself through a unique "Builder" layer approach. This approach enables low-code/no-code customization of pre-built applications, such as its Most Probable Cause and Log Analyzer applications, for specific customer environments using either large or small language models (LLMs/SLMs). The Acumen Builder capability operates at both the application level, for creating custom multi-vendor solutions, and at the AI services layer for chat and future agent capabilities. It supports both public cloud deployments via Amazon Bedrock and local AI services gateways for data sovereignty. Ribbon's approach leverages its Layer 0-7 expertise to provide real-time cross-domain visibility and automation. Early customer Optimum cited the platform's ability to enhance network reliability through predictive, AI-driven innovation and self-healing capabilities.

Ribbon also offers a flagship voice solution, Ribbon Analytics, which supports over 500 million subscribers and has been deployed in more than 100 instances. The solution is a big-data, cloud-native platform that employs machine learning for end-to-end network visibility, Quality of Experience (QoE) monitoring, advanced troubleshooting, and proactive voice threat prevention. Additionally, Ribbon's Muse Multilayer Automation Platform (MAP) for the IP Optical domain provides AI-enabled network management, including an AI agent that supports natural language queries. Their recently launched Acumen brings these platforms together and extends them with additional AI services and easier customization.

©2025 AvidThink LLC. All Rights Reserved.

Other Vendors

There were many other agentic solutions launched in 2025 by enterprise and carrier networking vendors. The constraint on report length limits how many we can cover, but we'll cite a few additional examples here: enterprise networking vendor, **Extreme Networks**, is integrating agentic AI, or intelligent agents, into its Extreme Platform ONE to automate network management, troubleshoot issues with greater accuracy, and streamline workflows. Startup **Aviz Networks** touts its Network Copilot as a private AI platform that provides operational assistance for data center and enterprise campus deployments. On the carrier side, operations support systems (OSS) vendors like **Amdocs** and **Netcracker** are rolling out agentic capabilities in their platforms. Meanwhile automation and orchestration players like **Network to Code** and **Itential** (with partner **Selector AI**) are offering AI-powered versions of their automation suites.

Barriers and Headwinds in 2025

While the pace of innovation over the last 12 months has been remarkable, the path to widespread adoption of advanced AI in networking is not without impediments. The barriers discussed in 2024 remain relevant, and the shift to "agentic" brings additional challenges.

The foremost concern we've heard from enterprises and carriers is **Trust and Control** in agentic systems. The prospect of AI agents autonomously executing changes on a production network requires careful risk management. The potential for "uncontrolled autonomy," where an agent makes an incorrect decision that leads to an outage, is a primary concern for operators. Early agentic deployments control for the "blast radius" — the scope of systems impacted if the autonomous agent fails, but over time, to reap the full benefits of agentic capabilities, the span of control needs to grow. This underscores the need for robust, explainable AI that can articulate its reasoning, as well as sophisticated governance and guardrail frameworks with multi-stage approval workflows, to ensure a human remains in the loop for critical actions.

A related challenge is the risk of **Agent Sprawl**. As enterprises and carriers begin to deploy AI agents, there is a significant risk of creating hundreds of siloed, custom-built agents for different tasks without a unifying architecture. This risk is amplified by the emergence of new workflow builders and agent canvases across multiple networking vendors. This could quickly lead to a new and insidious form of **Technical Debt**, creating operational chaos and making the system difficult to manage, monitor, and secure. The need for a cohesive "agentic AI mesh" or some centralized governing architecture to orchestrate these agents will be critical to long-term success.

Finally, foundational issues of **Data Governance** and **Privacy** persist and are even more crucial in the agentic era. AI agents require access to a vast and continuous stream of sensitive operational data, spanning multiple network domains and cloud environments. Ensuring that this data is handled securely, that privacy is maintained, and that data residency requirements are met — particularly when leveraging cloud-hosted AI models — remains a top-level concern.

The following 12 to 18 months will be defined by a race among the major platform players to prove the tangible ROI of their respective agentic frameworks.

The Human Element: Evolving Roles for the Agentic Era

The profound impact of AI on the roles and responsibilities of networking professionals, a topic of discussion in 2024, is now a present-day reality. The rapid shift toward intelligent, automated systems is starting to reshape careers, requiring a new set of skills to remain relevant even as AI systems shift from augmenting or assisting humans to being autonomous.

The role of the network engineer is evolving from a hands-on operator, proficient in vendor-specific CLIs, to that of an **AI Operator** or **Human Orchestrator**. The engineer's primary function will not be to configure individual devices manually, but to define high-level business intent and manage a fleet of AI agents that translate that intent into action. The focus shifts from the "how" of configuration to the "what" of desired outcomes, with the engineer's role becoming more strategic, analytical, and akin to a software developer managing an automated system. The other parallel roles for human engineers are as trainers and experts who help improve the models through **reinforcement learning**. By providing active feedback to models, human experts can help improve the performance of these models dramatically and accelerate the path to autonomous operations. We've seen early indications of this in the software development world, which is arguably the farthest along the autonomous agentic journey, with agents creating useful code, fixing bugs, and running tests with little human intervention.

©2025 AvidThink LLC. All Rights Reserved.

The ongoing evolution to autonomous networks necessitates a significant **upskilling** of the workforce. Proficiency in programming and scripting languages, such as Python, an understanding of APIs, familiarity with analytics, and expertise in major cloud platforms are essential for the modern network professional. The new, higher-level competencies in high demand include a working knowledge of data science and machine learning principles, as well as the ability to integrate and manage complex automation frameworks.

This transition also brings the **"Loss of Expertise" Paradox** into sharp focus. We're concerned that over-reliance on AI systems could lead to an erosion of the deep, hands-on expertise built over decades of manual troubleshooting. As a generation of veteran engineers retires, the industry faces a critical challenge: we must find effective ways to codify this invaluable, experience-based knowledge into AI models and training data before it is lost. Alternatively, we need to find ways to educate a new generation of network engineers.

Updated Recommendations for Enterprises and Service Providers

Navigating the complex and rapidly evolving landscape of AI in networking requires a deliberate and strategic approach. The recommendations from 2024 provided a solid foundation, but the emergence of "agentic" this year necessitates a more prescriptive and architecturally focused set of guidelines below:

- **Shift Evaluation from Features to Architectures:** The focus of procurement and technology evaluation must move beyond assessing a vendor's "AI assistant" or a specific AIOps feature. Instead, organizations should evaluate a vendor's entire agentic AI framework. This includes scrutinizing the underlying data model, the richness of the tools (APIs) available to the AI agents, fine-tuning capabilities to improve performance over time, and, most importantly, the robustness of the governance and oversight capabilities that allow for safe, controlled automation.

- **Prioritize the Data Foundation:** A vendor's AI is only as good as its data. Give strong preference to platforms that offer a unified, multi-domain data lake architecture. A platform that incorporates the inventory of all networking assets across the infrastructure and can also ingest, correlate, and store high-fidelity telemetry from across the network — both historically and in real-time — can provide the rich context essential for practical AI reasoning and accurate, proactive decision-making.

- **Embrace Multi-Vendor Orchestration:** Recognize that L4/L5 autonomy will be achieved in a step-wise, domain-by-domain fashion. As such, it is critical to invest in a multi-vendor automation and orchestration layer that sits above these autonomous domains. Find solutions that can manage end-to-end service lifecycles across multiple vendors' equipment and AI systems.

- **Invest Aggressively in Upskilling:** The skills gap between traditional networking and the new AI-driven approach is real and widening. Organizations must create formal, continuous training programs for their networking teams, focusing on Python, APIs, data analytics, ML, and automation principles. The role of the network engineer is now unequivocally a software-centric discipline.

- **Start with "Human-in/on-the-Loop" Agentic Use Cases:** Build trust in AI-driven automation by adopting a phased approach. Begin by deploying AI agents in an advisory or "copilot" mode, where they analyze situations and recommend actions that a human operator must review and approve. This "human-in-the-loop" model allows teams to validate the accuracy and safety of the AI's recommendations, build confidence in the system, and fine-tune its behavior before transitioning to more automated tasks. Once trust is established, transition to a "human-on-the-loop" approach, where human operators primarily monitor and intervene only when necessary. It may then be possible to go to autonomous, "auto-pilot" operations for well-understood, lower-risk tasks that require little oversight. However, beware of human operators who rubber-stamp AI recommendations without taking the time to evaluate — this "out-of-the-loop" (OOTL) behavior has been recognized and researched as part of autopilot and system automation for many decades[15]. OOTL can compromise reinforcement learning, and result in incorrect actions that have system outage ramifications.

[15] Endsley, Mica R.; Kiris, Esin O. (June 1995). "The Out-of-the-Loop Performance Problem and Level of Control in Automation". Human Factors: The Journal of the Human Factors and Ergonomics Society. 37 (2): 381–394. doi:10.1518/001872095779064555

©2025 AvidThink LLC. All Rights Reserved.

- **Don't Overlook the Cost (or Human) Element:** Reasoning models consume a large number of tokens, which can be costly. There are two strategies here that can help improve ROI. First, recognize that humans can and should continue to play a role. There is a productivity frontier in trading off agentic AI vs human intelligence, and the art is to balance between the two across a series of operational tasks to achieve good performance at the right cost. That balance will shift in favor of AI as computing costs decrease and models improve. Second, reduce computing costs by enhancing the agentic workflow to utilize smaller models (e.g., fine-tuning with distillation) and optimizing the system through various techniques (e.g., quantization, caching) to achieve increased efficiency. At the same time, recognize that processing costs will drop dramatically (up to 10X) every year with improved hardware, software, and model architectures. We recommend not over-investing in optimizing existing flows, but taking into account expected near-term cost improvements as workflows are designed and built.

Conclusion: The Path to 2026 and Beyond

The networking industry crossed a threshold in 2025. The transition from generative AI as a novel interface to agentic AI as a core architectural principle is the most crucial development in network automation since the advent of software-defined networking (SDN) in the early 2010s. The arrival of agentic capability finally unlocks the path towards intent-based networking.

The vision of a fully autonomous network, long considered a distant "pipe dream," is now on a clear, albeit complex, path to becoming a reality. This reality will not arrive as a single, monolithic event but as a mosaic of autonomous domains, each managed by specialized AI systems. The following 12 to 18 months will be defined by a race among the major platform players to prove the tangible ROI of their respective agentic frameworks.

Simultaneously, domain-specific vendors will continue to deepen their domain-specific autonomous capabilities, offering best-of-breed solutions in specific segments of the network. For enterprises and service providers, the defining challenge and opportunity will be to develop the architectural and human capabilities needed to manage this new, hybrid-intelligence, multi-vendor world. The journey is underway, and the pipe dreams of yesterday are becoming tomorrow's operational realities.

We hope you've found the content in this research brief helpful and welcome your input and feedback. You can reach our team at research@avidthink.com.

SPONSOR INTERVIEWS SECTION FOLLOWS

Please support our sponsors — check out in-depth interviews with their experts on AI in Networking. We encourage you to visit their websites to learn more about their solutions.

©2025 AvidThink LLC. All Rights Reserved.

Interview with Ram Ramanathan, Vice President, Product Management

The networking industry has been talking about autonomous networks for decades, yet we're still far from that reality. What's your perspective on where AI fits into this journey today?

We're seeing AI deliver real value in specific domains, particularly around fault management and network optimization, which consistently rank as high-value outcomes for carriers. The key is to recognize that autonomous networks won't emerge overnight — it's an evolutionary process. What's changed is our ability to apply AI not just to detect problems, but to understand complex multi-vendor environments and provide actionable insights across the entire organization, from NOC operations to customer experience management.

Networking vendors talk about AI-driven automation, but customers still face complex multi-vendor environments that resist traditional automation approaches. How is Ribbon addressing this challenge differently?

The key differentiator is our low-code/ no-code solution element, Acumen Builder. While most vendors offer pre-canned AI applications for fault detection or log analysis, we customize these to each customer's specific multi-vendor environment. For example, we're working with a customer that has Juniper routers, Cisco and Arris cable systems, and Arris set-top boxes. Rather than treating this as a generic network, Builder makes it easy to add customization layers that adapt our pre-built applications—such as Most Probable Cause and Log Analyzer—to then understand the unique behaviors, configurations, and interactions of that specific vendor mix. What we're doing isn't just data customization, but logic and model customization to deliver precise insights for their environment.

Many carriers struggle with operationalizing AI beyond pilot projects. What's preventing broader deployment, and how do you help customers move from proof-of-concept to production?

The most significant barrier is the gap between having data and having actionable insights across organizational silos. We've seen customers where NOC teams, customer support, marketing, and business units all need different views of the same network intelligence. As an example, one of our customers regularly has 1,000 users logged on, from various departments, not just network operations. The challenge isn't just technical; it's organizational. We help bridge this by providing role-specific dashboards and interfaces that translate network insights into business value, whether that's reducing customer churn or optimizing marketing campaigns based on service quality metrics.

Your approach emphasizes collaboration across different teams within telecom organizations. Can you elaborate on how this differentiates Ribbon's go-to-market strategy?

What we've discovered is that many carriers have talented data scientists and in-house AI capabilities, but they struggle with operationalizing these insights across domains. Ribbon's unique DNA - spanning layer 0 to 7 with expertise in IP, optical, cloud-native platforms, and edge technologies - allows us to speak the language of every team that needs to collaborate for successful AI implementation. We recently conducted an automation workshop with 28 attendees from various departments at one customer. The magic happened when the NOC team's operational insights connected with marketing's customer churn data. We're not only providing technology; we're facilitating the organizational transformation needed to capture AI's full value.

Looking ahead, do you see AI automation working with vendor-specific intelligent agents, or will multi-vendor platforms like yours remain necessary?

Long-term, we anticipate a hybrid model. Vendor-specific agents will emerge - say a Cisco smart agent or Juniper intelligent system. Ribbon's role will evolve to orchestrate various agents through standardized protocols, such as MCP or A2A. Think of it as conductor-level intelligence that coordinates multiple vendor agents to deliver end-to-end automation. In the near term, however, most networks remain multi-vendor environments where our generic approach to data analysis provides immediate value. In the mid to long term, the industry will work toward agent interoperability standards.

blueplanet
a division of Ciena

An Interview with Jean-Philippe Goyet
Sr. Director of Product Management, AI

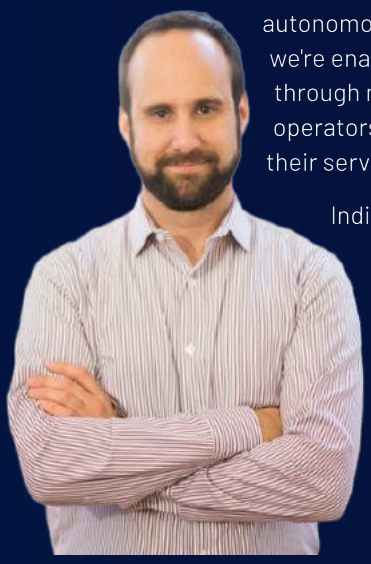
We've seen a lot of hype around AI in networking. What's Ciena Blue Planet's practical approach to implementing AI and generative AI in network operations?

Our strategy is focused on improving and defining new capabilities within our existing applications rather than building AI for AI's sake. We're taking a dual-pronged approach: first, we're building out-of-the-box AI applications and agents that are consumed by our different network management applications. Second, we're enabling our customers and partners to enrich these capabilities through our platform.

We announced at TMForum Digital Transformation World that we're extending our AI Studio product to an agentic framework with three distinct layers. The foundational layer is the agentic core - the ability to build, run, manage, and orchestrate AI agents across applications. The second layer provides agentic tooling that gives agents access to all the rich data and APIs across our OSS use cases. The third layer is how applications actually consume these agent capabilities for inventory, orchestration, assurance, and network slicing.

Can you give us some concrete examples of how these AI agents are transforming network operations today?

Let me share a specific example we're working on with 5G network slicing. There are two key aspects we're improving: first, the system's ability to better understand natural language and abstract user intent, then convert that into actual network resources. This is a major autonomous networking use case. Second, we're enabling the creation of new services through natural language - allowing operators to create new products within their service catalog dynamically.



Individually, these use cases bring significant efficiency improvements to current processes. And when combined, they offer a completely new way to provide dynamic offerings to end customers. We're optimizing existing processes while transforming the types of products that can be offered.

Network planning has always been complex and time-consuming. How is AI changing this process?

Planning is where everything starts for carriers, so it's crucial to get right. We're leveraging agents to interact with end users in new ways, combined with graph neural network technology that predicts the impact of network changes. Whether you're adding capacity, changing routing, or adding new links or routers, our system can predict outcomes before implementation.

We're also using generative AI for configuration automation. When deploying complex systems like Blue Planet, there's extensive configuration work required for specific implementations. AI can automate device model configuration in inventory and eliminate manual processes. The goal is to transform the user experience into a more conversational approach for accessing data while enabling an ecosystem of external agents that might need knowledge from our systems.

Given that most carrier networks are complex, multi-vendor environments that have evolved over decades, how realistic is the timeline for AI transformation?

You're right that this is a massive, complex field. When you work in OSS environments, you're dealing with all network layers, multiple domains, and most carriers worldwide use multiple equipment vendors. We've spent years on standardization just to make these elements communicate effectively.

The change won't happen overnight. I learned from my experience with SDN and network virtualization that things progress gradually. While the technology was exciting in 2015, it took years to actually move networks to those new technologies. AI and agents will follow the same pattern.

I see this as a progressive transformation starting with interfaces - how service catalogs are exposed to BSS systems or portals. Then we'll see changes in integration with controllers and orchestration systems. We need to understand which processes have to be deterministic versus which can leverage probabilistic reasoning using agents. This will drive progressive transformation of the entire stack as agent technology also evolves, ultimately changing the fabric of network orchestration while transforming internal workflows.

Q&A with Wim Hendrickx, CTO, IP Division

Nokia has communicated goals to achieve level four autonomy as defined by the TM Forum framework. What does this mean practically for network operators, and how do you aim to tackle this huge problem space?

We're assessing our entire product portfolio against TM Forum's operational flows. Some capabilities are already at level four, others are progressing. Over time, operators will see more capabilities helping them achieve autonomous network operations.

The problem space is enormous, but I see us taking a divide-and-conquer approach to make it more manageable. We're breaking down operational flows into smaller domains – mobile, fixed, optics, IP – and building expert systems around each. Predicting optical failures requires different expertise than routing or control plane issues. These specialized expert systems collaborate to achieve higher-level autonomous operations.

Q: Network environments typically involve multiple vendors, each with their own AI approach. How does Nokia address this fragmentation, and what's your vision for industry standardization?

This is a significant challenge since every vendor believes they can build the best solution. The reality is customers have multiple vendors and need unified consumption models rather than isolated building blocks.

Our approach focuses on well-defined interfaces to expert systems. We need industry agreement on protocols for communicating with vendor-specific expert systems. Initiatives like Model Context Protocol (MCP) are good examples. Once we establish these interfaces, we can build collaborative systems where expert elements communicate to achieve bigger goals. Even within Nokia, different groups contribute different expertise, so we're building systems that easily integrate together.

You've mentioned that the solution isn't AI-only but requires a hybrid approach including agents and human oversight. Can you provide a concrete example?

Today we use humans and programs together. Now we're adding AI systems as a third element. It's about choosing the right tool for each task.

Take our path computation element for traffic engineering and optimization – it's hardened, works at scale, and proven with customers. There's limited benefit moving that to AI. However, AI excels at troubleshooting assistance. Even there, combinations work best. In an MCP approach, LLM models are assisted by tools, which are typically programs. Our assessment shows optimal solutions combine AI technologies, ML, programs, and human assistance collaborating for specific operational tasks.

Trust is critical for autonomous systems. How do you build confidence in AI-driven network operations?

We employ multiple validation techniques. In particular, domain controllers play a crucial role. They provide validation through digital twins and other logic, curating answers and providing confidence scores. We validate through programs and multiple AI approaches together, then curate results to ensure we're heading in the right direction. This ensures proposed changes will actually work before implementation.

With fewer network engineers entering the field and existing talent aging out, how does Nokia's AI strategy address this skills shortage while preserving critical expertise?

We're taking a three-pronged approach. First, we're making our products and documentation more easily consumable through natural language interfaces, so you don't need to be an expert to understand them.

Second, we're offering consultative services to transfer our in-house expertise to customers building networks. Third, we're developing SaaS-based offers for smaller customers who want our technology but lack operational skills. We handle upgrades, patches, and capabilities as a service.

We're even moving toward proactive issue identification using knowledge bases combined with AI and LLMs. Our goal is identifying and troubleshooting issues before customers notice them, essentially becoming partners in their operations.

Q&A

Kishan Ramaswamy
Director, Product Management
(VeloCloud - acquired by
Arista Networks)

Brendan Gibbs
Vice President, AI, Routing, and
Switching Platforms, Arista Networks

VeloCloud has observed significant AI application deployment across enterprise networks. What are you seeing in terms of bandwidth consumption and traffic patterns from these AI workloads?

Kishan: We conducted an analysis across more than 200 customers with at least a thousands sites each, and the results are striking. Overwhelmingly, across all verticals, we found that there was significant uptick in the number of AI applications accessed. These have been , predominantly GenAI applications. This creates unique network requirements that enterprises must address.

AI applications have distinct characteristics that networks need to accommodate. First, there's an expectation of fast response times – when users send queries to AI applications, they expect equally fast responses. This demands low-latency network capabilities, which VeloCloud's overlay technology can provide when the underlying network falls short.

Second, these applications consume substantial bandwidth with symmetric traffic patterns. Gone are the days of asymmetric bandwidth with large downloads and small uploads. AI applications send significant amounts of data for processing and receive comparable volumes back. Finally, the traffic is extremely bursty – you don't see steady-state AI queries, but rather bursts when users need something. VeloCloud is specifically built to address these unique traffic characteristics.

How is VeloCloud leveraging AI to improve network operations and troubleshooting for your customers?

Kishan: VeloCloud incorporates AI across two primary areas. First, our AIOps functionality provides intelligent network troubleshooting. When alerts appear on the orchestrator – like a link going down or flapping – users can click on that alert and leverage our AI analytics to understand why it happened. The system analyzes what occurred prior to the fault, suggests corrective measures, and can even automatically create support cases with Arista. This drastically reduces the time network engineers spend troubleshooting issues.

Second, we use AI and ML for application identification. Many applications today are encrypted from end to end, making it challenging to know what application you're dealing with. VeloCloud uses underlying AI technologies to identify applications based on their behavior patterns, flow characteristics, and setup parameters. Only by accurately identifying applications can you enforce appropriate policies, which is crucial for optimal network performance.

We're also developing agentic AI capabilities that will enable operations teams and network engineers to interact with intelligent agents for streamlined problem resolution and root cause analysis.

Now that VeloCloud is part of Arista, how will the combined companies leverage AI across the broader network infrastructure?

Brendan: We're incredibly excited about the possibilities. As a newly merged company, we're looking to gain AI insights across a much broader range of streaming telemetry – from wired and wireless solutions in branch offices to SD-WAN and Cloud-WAN connectivity back to headquarters. We see utilizing the added telemetry into the Arista CloudVision platform to bring improved end-to-end observability and value to our customers.

Our goal is seamless cross-domain visibility and AIOps capabilities that span from the branch office all the way back to headquarters. By combining Arista's switching and wireless expertise with VeloCloud's SD-WAN capabilities, customers will get enhanced AI value from traffic flowing across the entire infrastructure and more actionable AI outcomes across our the breadth of our solutions.

Q&A with Benjamin Hickey, Director, Portfolio Management and M&A at IBM Software Networking

The networking industry has been talking about autonomous operations for years, but we're still largely in manual mode. What's IBM's perspective on why we haven't achieved true network autonomy yet?

Networks are different from the rest of the IT stack in two critical ways: scale and impact. Networks suffer from what we call the "shared fate problem" — they're distributed systems where the management scope is exponentially larger than other IT elements. And when networks fail, they have the potential to impact services and applications in the most devastating way. We believe the path to autonomy requires thinking about networks like our brains — you need both analytical AI for the left-brain functions and generative AI for the right-brain creative reasoning. Most AI innovation today focuses only on the GenAI right-brain.

You mention that current GenAI innovations are primarily "right brain" capabilities. What's missing from the "left brain" that's hindering true network intelligence?

The analytical challenge is enormous. Networks generate massive volumes of telemetry data, and we need to dramatically improve our signal-to-noise ratio. LLMs can't understand time-series data or temporal relationships the way networks require. We need specialized time-series foundation models that can understand the nuanced behavior of specific network domains – whether that's IP, SD-WAN, data center, or radio access networks. Our approach uses ultra-compact models like the IBM Time Series Pulse with just 1M parameters, specifically pre-trained for network domains, delivering 1,000x improvement in latency with millisecond inference times. This approach delivers more accurate observations, finds issues otherwise missed, and provides early warning for brownouts.

How does IBM Network Intelligence address the challenge of understanding what "normal" looks like in each unique network environment?

This is where our architecture is unique. We analyze live data, with continuous baselining in the pipeline to handle the unique personality of every metric in your network. It determines whether a metric is highly predictable, seasonal, noisy, or erratic. This process utilises multiple models in our pipeline to get the results that have not been possible with purely statistical analysis with ML algorithms. It's about having the right tool for the right job at the right cost, ensuring speed and relevance before data reaches our more sophisticated foundation models.

Network operations teams are drowning in alerts and false positives. How do you move from raw observations to actionable intelligence?

Raw observations lack context. Knowing there's high latency on a link doesn't tell you if it's on a critical path, which applications depend on it, or whether this has happened before. Our agentic right-brain reasoning takes observations utilizing the domain knowledge we have trained it on to develop root cause hypotheses. These are run through critical filters with contextual proprietary documents such as operators' knowledge bases, and can use external systems to search tickets, and query APIs and network sources of truth for real-time context. Our reasoning LLM then forms remediation plans, again based on the operator's proprietary operating guides and procedures. This iterative loop of hypothesis generation, data validation, and bespoke remediation planning enables deep contextual reasoning with the trust and explainability operators need.

Trust seems to be a major barrier to AI adoption in network operations. How are you addressing this challenge?

Trust is absolutely critical, which is why we've built extensive scaffolding around our LLMs – think of it as guardrails for the AI's thought process. We use standardized prompts and tool guardrails to make the investigative process consistent, repeatable, and safe. We even have a context-locked assistant that only answers network-relevant questions. Our vision is of a system that is designed to learn through reinforcement learning when operators accept hypotheses, and to have subject matter experts be able to directly refine the agent's reasoning traces, teaching it better investigation methods. This means the system becomes progressively smarter and more effective over time.

ERICSSON

Interview with Archana Khetan

SVP, Product Management and Technical Marketing
Ericsson Enterprise Wireless Solutions

How has Ericsson's AI journey evolved in networking operations, and what role does your virtual assistant play in this transformation?

Our AI journey at Ericsson has accelerated significantly over the past year. Initially, we were primarily focused on AIOps, especially in SD-WAN and SASE environments. Late last year, we introduced generative capabilities into our AI-based NetCloud Assistant (ANA), which brought guided documentation access and personalized recommendations to our customers via a RAG framework. ANA now provides more thoughtful answers with contextual information gathered from our NetCloud Manager platform, as well as technical documentation. This year has been particularly exciting as we've entered the realm of agentic AI. While LLMs can reason and answer questions, when you want orchestration of complex tasks and broader planning and analysis - especially automation of activities - that's where the agentic framework becomes critical.

Where do you see AI adding the most value across the network lifecycle, from Day 0 design through Day 2 operations?

We conducted a survey of support cases across our installed base and found that troubleshooting represented 25% of cases, but 75% were around "how to" questions and product-level information. This validated that ANA's guided documentation capabilities were addressing a main pain point by enabling customer self-service. Our immediate focus is on Day 2 operations since that's where our mature customers spend most of their time. Later this year, we'll be launching our first orchestrator agent capability focused on troubleshooting. Moving into 2026, we'll expand into policy, configuration, and deployment elements covering Day 0 and Day 1 activities.

What challenges are you experiencing with AI implementation, and how are you addressing issues like explainability and trust?

We are addressing two significant infrastructure-level challenges. First, a good AI strategy requires a robust data strategy. We're doing significant internal work to ensure all datasets that customers could potentially use in their AI journey are meaningfully available to our AI systems. Second is explainability of AI, which is a common industry problem. This quarter, we're launching enhanced ANA capabilities that show customers in real-time what actions ANA is taking and what's happening in the background. This transparency ensures that what comes out of our virtual assistant isn't a black box, but provides a greater understanding of ANA's internal processes. Additionally, all of our models are securely hosted within Ericsson's environment, which means data is not being transmitted externally for processing. On top of that, data including personally identifiable information is anonymized using one way hash before it is stored in the data lake for model training. These elements — building customer trust in AI and getting meaningful data to solve the problems we're targeting — are our most important focus areas.

How does Ericsson's approach to AI in networking differentiate from other SD-WAN and SASE providers?

Our differentiation lies in our focus on 5G intelligence. While many SD-WAN vendors base their policies on generic requirements that are similar across the industry, our strength lies in providing 5G intelligence that makes SD-WAN policies much more targeted at delivering the right quality of experience wherever 5G, cellular, or satellite connections are involved. This isn't just a technology play — it has business implications. Our policy decision-making goes deeper into commercial elements, including pooling of data plans and pricing availability, to determine which link is most economically viable while still meeting quality of experience requirements. These 5G differentiation elements carry forward into all AI-powered policy decisions.

Looking ahead, what's your vision for how AI will transform the customer experience and network operations for enterprises?

Our vision is that NetCloud Manager needs to evolve so every config page and every layer is AI-enabled. We want customers to recognize the vision of fully autonomous, self-healing networks that we've all been aspiring to achieve. Networks are becoming increasingly complex, and our vision centers on intent-based requirements, with AI handling the complex orchestration and configuration to meet those intents. This addresses the critical talent shortage we're seeing. When I visited European customers last year, the lack of expertise was delaying deployments because the equipment was sitting in warehouses waiting for availability of experts. I'm looking forward to helping our customers overcome network complexity and the skills shortage by making it easy for network administrators to configure environments through intent-based prompts and by providing the intelligence to increase the reliability and quality of experience of 5G connections.

AvidThink, LLC
1900 Camden Ave
San Jose, California 95124 USA
avidthink.com

©2025 AvidThink LLC. All Rights Reserved.
This material may not be copied, reproduced, or modified in whole or in part for any purpose except with express written permission from an authorized representative of AvidThink LLC. No part of this work may be used or reproduced in any manner for the purpose of training artificial intelligence technologies or systems. In addition to such written permission to copy, reproduce, or modify this document in whole or part, an acknowledgment of the authors of the document and all applicable portions of the copyright notice must be clearly referenced.

Rev A