

Pipe Dreams and AI Realities – Networking’s Midlife Crisis

Examining the Promise of AI in Networking in 2024

RESEARCH BRIEF



Table of Contents

- Executive Summary..... 1**
- Introduction – Addressing AI Hype and Reality in Networking2**
- AI is Critical in Realizing Autonomous Networking2**
 - Autonomous Network Maturity model.....3
- AI for Autonomous Networking – Predictive and Generative Models 4**
 - Foundation Models and their Role in AIOps4
 - Market Surveys on GenAI Adoption in Networking5
 - Role of AI Agents6
 - Natural Language Processing.....7
 - Intermediate Representations – Digital Twins and Network Models.....7
- AI Capabilities for Network Operations8**
 - Discovery8
 - Design and Configuration Generation9
 - Automation9
 - Prediction9
 - Simulation.....9
- AI Impact Across the Network Lifecycle9**
 - 1. AI-Assisted Network Design (Day 0)9
 - 2. Automated Deployments (Day 1) 10
 - 3. Management and Monitoring (Day 2)..... 11
 - 4. Optimized Environments 12
 - 5. Troubleshooting and Customer Support..... 13
 - 6. Continuous Improvements..... 14
- Use Case Mapping to Expected Solution Maturity 15**
- Real-World Use Cases of AI in Networking 15**
- Networking Vendor AI Implementations..... 16**
- Barriers to AI Adoption 18**
- Impact of AI on People and Organizations 19**
- Recommendations for Enterprises and Carriers 19**
- Wrapping Up 20**

Pipe Dreams and AI Realities – Networking’s Midlife Crisis

Examining the Promise of AI in Networking in 2024

Executive Summary

The networking industry has dreamed of autonomous networks for many years – self-operating, self-healing intelligent pipes that connect businesses and consumers globally. However, decades after the launch of commercial carrier and enterprise networks, fully autonomous networks remain but a vision. Nevertheless, AI is changing how networks are built and operated, from design and deployment to management, optimization, and continuous improvement. Traditional predictive AI/Machine Learning (ML) and emerging Generative AI (GenAI) technologies are delivering benefits across the network lifecycle.

Predictive AI has proven its value in anomaly detection, capacity planning, and energy optimization. For example, telecom operators have reduced power consumption in mobile networks using AI-powered solutions, and AI helps optimize enterprise Wi-Fi radio resource management. Meanwhile, GenAI shows promise in enhancing human-computer interfaces, assisting with complex troubleshooting, and enabling more intuitive network configuration. Companies like eBay, Orange, and Google leverage AI to predict potential network failures, automate trouble ticket generation, and streamline incident analysis, reducing problem resolution times and improving network reliability.

Despite these successes, barriers to widespread AI adoption in networking persist. These include concerns about trust in AI-driven decisions, challenges in scaling early GenAI pipeline architectures like Retrieval Augmented Generation (RAG) approaches, the need for explainable AI, data privacy issues, and the complexities of implementing AI in fragmented, multi-vendor network environments. Additionally, the organizational impact of AI adoption, including potential changes to workforce roles and skills, requires careful management to ensure smooth integration and maximize benefits.

While navigating this complex landscape, enterprises and carriers should adopt a measured, phased approach to AI implementation. Users should start with proven AI/ML solutions, focusing on use cases with clear ROI, investing in robust data infrastructure, and embracing digital twin technology for risk-free simulations. Maintaining human oversight, including adopting AI-copilot approaches, remains crucial while AI capabilities grow. By balancing innovation with caution, organizations can harness the power of AI to enhance network efficiency, reliability, and user experience while preparing for the future of autonomous networking.

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, SSE, ZTNA, cloud 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.

Introduction – Addressing AI Hype and Reality in Networking

The current wave of generative AI (GenAI) hype has grabbed global attention, becoming a board-level topic for networking vendors and operators. However, artificial intelligence (AI), including machine learning (ML) and deep learning (DL) techniques, has been used in networking for over a decade – even though their impact may have been invisible. From improving radio resource management in Wi-Fi networks to optimizing spectrum use in mobile radio access networks (RAN) and aiding backbone capacity planning, traditional predictive AI/ML has significantly benefited the networking field.

Despite over two decades of vision and dreams, along with research and investment, fully autonomous networks remain an unattainable goal for now. As carrier and enterprise networking enter the middle phase of their commercial lives, the resurgence of interest in AI/ML has ignited a new wave of innovation – similar to the impact of software-defined networking (SDN) a decade ago. The arrival of OpenAI's ChatGPT triggered a scramble by carriers, enterprises, and networking vendors to apply GenAI to their businesses, with mixed success to date.

This report aims to surface near-term value companies can create and capture using AI/ML and GenAI in network design, automation, and operations – sometimes termed Networking AIOps (AI operations). Based on extensive conversations with networking vendors, researchers, and practitioners, our report captures:

- a) The role AI can play in delivering value on the industry journey to autonomous networking
- b) A mapping of use cases to the maturity model
- d) Select real-world examples showcasing AI/ML in networking
- c) Sample capabilities available from vendors today
- d) Predictions for future evolution
- e) Recommendations for networking executives and technology leaders

Feedback is always welcome at research@avidthink.com.

AI is Critical in Realizing Autonomous Networking

The concept of an autonomous network – one that operates independently with minimal human oversight – is not new. Early ideas of self-organizing, self-optimizing, and self-healing networks emerged in research in the 2000s, and standards bodies like 3GPP highlighted them circa 2008/2009. Since then, the concepts have continued progressing across carrier and enterprise networks. The growing scale and complexity of today's enterprise, data center, and carrier networks, coupled with the criticality of our communications infrastructure, makes it imperative to redouble efforts in achieving higher levels of autonomy.

Unlike traditional networks that operate on predefined rules with limited ability to learn or adapt dynamically, autonomous networks can self-configure, self-manage, and self-heal, diagnosing and fixing problems as they arise. These networks can dynamically manage and optimize resources, redirect traffic flows for improved user experience, and enhance security measures without direct human intervention. The practical applications and cost-reduction implications of autonomous networks are far-reaching, spanning cloud data centers, campuses, wide area networks, 5G RAN, and industrial IoT.

At their core, autonomous networks are built on foundational elements that include:

- **Control:** Network virtualization, programmatic controls (APIs), and software-defined networking (SDN) capabilities that allow for agile management and efficient allocation of network resources.
- **Observability:** Data analytics and telemetry provide a continuous stream of data that enables intelligent decisions, allowing the network to anticipate and respond to potential issues reactively and proactively. In some implementations, Digital Twins are used as a proxy that reflects network objects' state (and history) and aids with convenience, scale, disconnected operations, and simulations.
- **Intelligence:** AI and ML provide the brains, enabling the network to make intelligent decisions and adapt to changing conditions in real time.

Recent advancements in AI have the potential to transform how networks are designed, configured, managed, optimized, and secured with minimal human intervention. AI, ML and potentially generative AI bring this vision closer to reality.

Autonomous Network Maturity model

The TM Forum’s maturity model can be a helpful guide to measure the evolution of autonomous networks from level 0 to level 5 (inspired by the SAE Levels of Driving Automation). While typically used by telecommunication carriers, it is domain-agnostic and equally applicable to enterprise networks. The six-step model outlines the journey from basic automation of repetitive tasks to achieving a self-aware and self-governing operational model, addressing the challenges faced at each step.

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

- **Level 0 – Manual Management:** The system delivers assisted monitoring capabilities; all dynamic tasks must be executed manually.
- **Level 1 – Assisted Management:** The system executes specific repetitive sub-tasks based on pre-configuration to increase execution efficiency.
- **Level 2 – Partial Autonomous Networks:** The system enables partial automatic O&M for specific units based on predefined rules/policies in certain external environments.
- **Level 3 – Conditional Autonomous Networks:** Building on L2 capabilities, the system with awareness can sense real-time environmental changes and, in some network domains, optimize and adjust itself to the external environment.
- **Level 4 – High Autonomous Networks:** Building on L3 capabilities, the system enables analysis and decision-making in a more complicated cross-domain environment based on predictive or active closed-loop management of service and customer experience-driven networks.
- **Level 5 – Full Autonomous Networks:** This level is the goal for telecom network evolution. The system possesses closed-loop automation capabilities across multiple services, domains, and the entire lifecycle, achieving autonomous networks.

AI for Autonomous Networking – Predictive and Generative Models

While the rise of GenAI has made AI more universally accessible, it doesn’t render existing Predictive AI strategies obsolete – GenAI and Predictive AI approaches complement each other well. Predictive AI helps in anticipating and preparing for future events, while GenAI enables innovation and exploration of new solutions in networking. Along with Agents, Natural Language Processing (NLP) systems, use of Digital Twins, and abstraction models, these collective techniques offer a comprehensive toolkit for building autonomous network systems.

	Generative AI	Predictive AI
USES	Create content – text, images, audio, video	Forecast events, classification
TECHNIQUES	Variational auto-encoders, transformers, generative adversarial networks (GAN), diffusion networks	Linear regression, decision trees, recurrent neural networks (RNNs) including long short-term memory (LSTM), convolutional neural networks (CNNs)
DATA INPUT	Large text corpora, images, videos	Historical data and events
COMMON APPLICATIONS	Text generation, image creation, music composition, new protein discovery	Time-series forecasting - stock prices, network utilization, demand forecasting
EVALUATION METRICS	Creativity, relevance, accuracy	Accuracy, precision, recall
CHALLENGES	Hallucinations, biases, computational cost, explainability	Data quality, overfitting, interpretability

Predictive AI, which has been deployed in networks for some time now, focuses on forecasting future events based on past data. It relies on algorithms such as linear regression, decision trees, and ML models like recurrent neural networks (RNNs) to identify time-varying patterns and predict what might happen next. The vision of autonomous networking leans on predictive AI for Day 2 tasks in network optimization, anticipating network failures before they occur, predicting peak times, and adapting capacity to changing network loads.

Generative AI focuses on leveraging the emergent capabilities of foundation models (FMs), including large language models (LLMs) and multi-modal models, to help with tasks in human-computer interfaces (HCI) and planning and “reasoning.” Many of the FMs used in GenAI are trained on large corpora of information from both open and closed sources and show remarkable communication, coding, and data analysis capabilities. Early indications suggest that taking a pre-trained FM and fine-tuning it with networking domain-specific datasets can drive innovation in numerous areas, such as AI-assisted chatbots, assistive network configuration, network scenario simulation, and root cause suggestions during troubleshooting. Additionally, GenAI can help infer human intent during network configuration, suggest new network designs that optimize for cost and performance, or simulate cyber-attacks to test network security.

Foundation Models and their Role in AIOps

Foundation models are at the heart of why generative AI has captured consumer and business interest. The FMs that power chatbots and the creation of text, images, music, and video content have demonstrated increasingly impressive capabilities. Pre-trained LLMs demonstrate robust out-of-the-box summarization and writing prowess, along with emergent capabilities in tasks that require planning, logical reasoning, and understanding of human intent and sentiment. While debate continues around the “intelligence” of these LLMs and FMs, and efforts are underway to combat hallucinations, toxicity, bias, and less desirable traits, these models have shown real-world value across multiple tasks.

Today, several commercial and open (open weights, open source, transparent training data, etc.) FMs are available, and more are being developed that can be used for GenAI in network operations. However, generative AI is early in its evolution, and many questions remain unanswered:

- Will the leading transformer architecture (and its variants) continue to scale?
- What happens when we run out of freely available data? How far will using synthetic data to train larger models take us?
- What’s the proper model pipeline for different network operation tasks? Large pre-trained models with prompt engineering and data augmentation? Or will smaller networking-specific custom-trained models perform better?

- Where will the energy to feed all the computation come from?
- Is the current rate of investment in model training sustainable without corresponding near-term returns on investment?
- Can GenAI models with appropriate workflows and guardrails achieve the reliability needed for fully automated networks without human intervention?

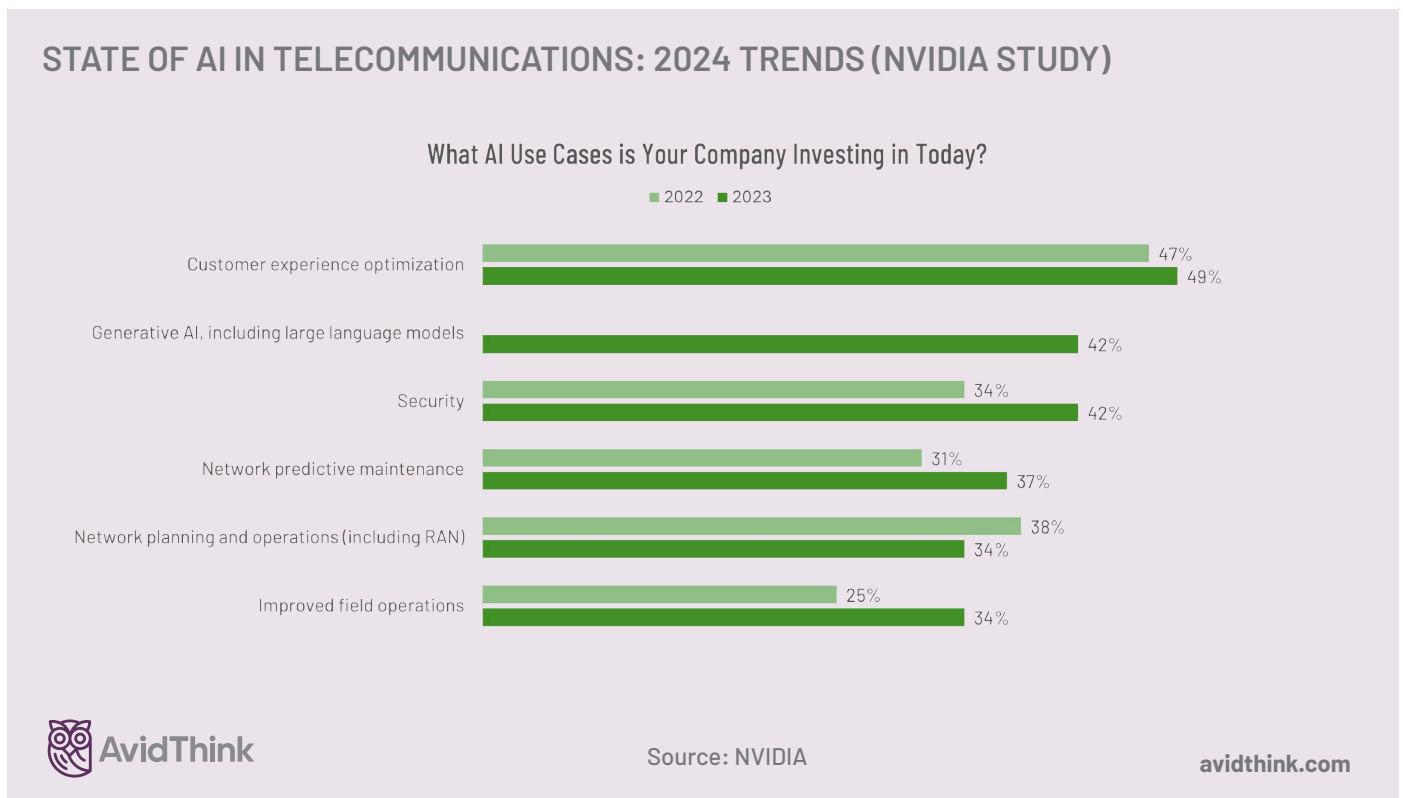
As of this writing, the top LLM models enterprises and vendors have leveraged for integrated AI capabilities in the networking domain include:

- Llama by Meta (Facebook): A popular open-weight choice, Llama is noted for its performance, even compared to models with more parameters from other organizations.
- Mistral: Another open-weight option, Mistral distinguishes itself with a strong emphasis on efficiency and versatility. It delivers high-quality outputs while maintaining lower computational costs. Mistral also focuses on modularity, allowing easy integration and customization to suit various industry-specific needs.
- GPT-4x by OpenAI: While proprietary and closed, the GPT-4 family continues to be a leader in the field, acknowledged for its advanced capabilities in understanding and generating natural language.
- Claude-3+ by Anthropic: Another proprietary option, Claude has gained attention for its enterprise-centricity, performance, and approach to AI safety, emphasizing helpful, honest, and harmless interactions.

The market for generative AI is early, and we expect vendor implementations to change as models evolve rapidly. In general, human-computer interface approaches can be supported by smaller models. Further fine-tuning for the networking domain with networking concepts and terms can improve their performance without a much larger footprint. However, larger FMs might perform better for reasoning tasks like troubleshooting, fault isolation, and root cause analysis.

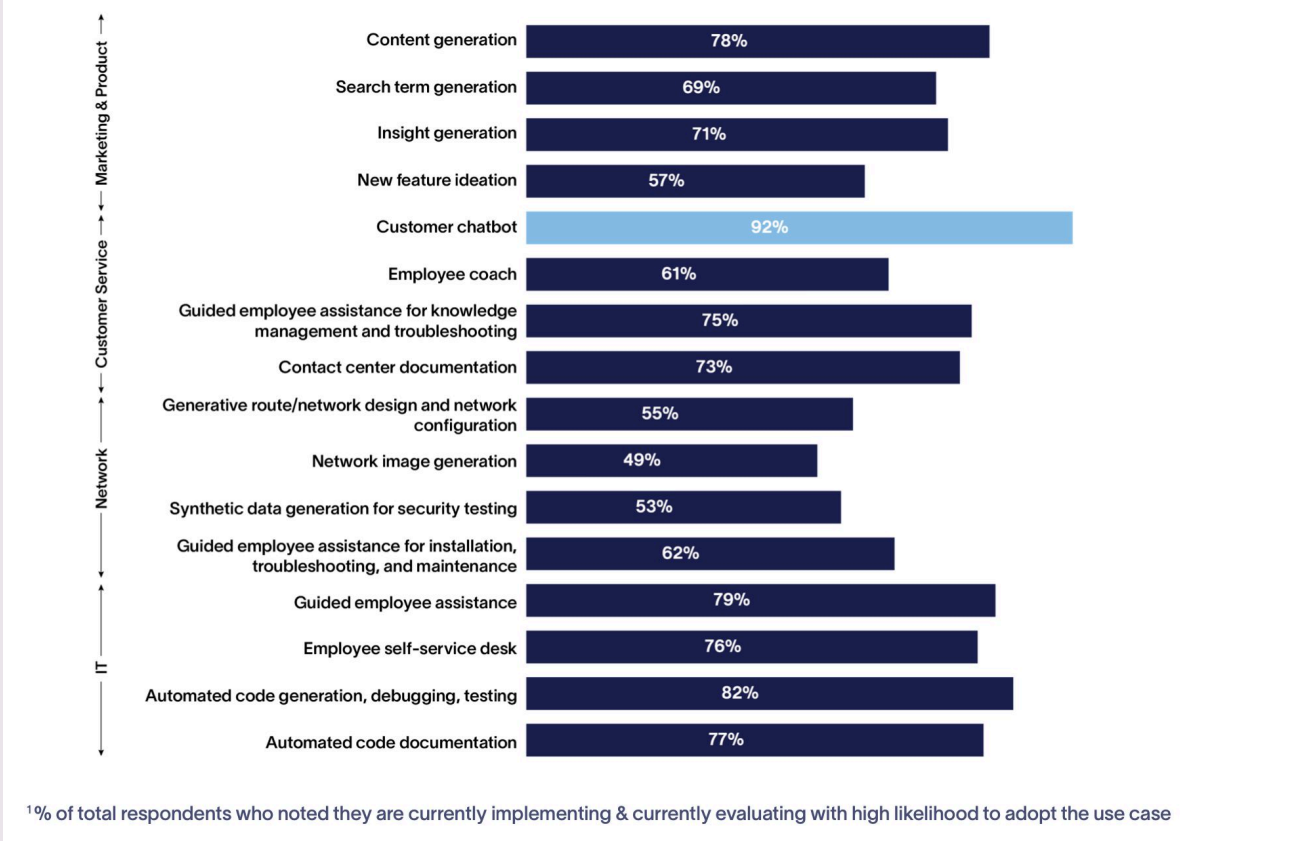
Market Surveys on GenAI Adoption in Networking

As evidence of the immaturity of GenAI use in networking, here are snippets from two recent surveys covering generative AI's adoption amongst CSPs (one performed by consultancy Altman Solon and commissioned by AWS, the other from NVIDIA).



GENERATIVE AI TELECOMMUNICATIONS-SPECIFIC USE CASES

% of total respondents, N varies by respective domain



Source: Altman Solon

avidthink.com

The 2024 NVIDIA study¹ shows that AI use in field operations, network planning and operations, and predictive maintenance will receive less investment from telcos than customer interactions (support, self-service, etc.). Both reveal that while GenAI-powered customer chatbots are now better understood and more widely piloted and deployed, GenAI adoption in network operations lags behind other use cases. Similarly, the Altman Solon survey² commissioned by AWS shows customer chatbot and marketing use cases being more popular than guided troubleshooting and network design and configuration.

Role of AI Agents

In the broader GenAI ecosystem, there are discussions that the next stage of evolution is using “agents,” or where FMs are used in agentic tasks. In network automation and operations, the term “agents” is not novel, but the definition is evolving, and network agent capabilities are expanding. Typically, network agents are software components or bots that can perform predefined actions based on rules, triggers, or inputs. Acting as the execution engine for automated tasks and workflows, they play a crucial role in network automation. An agent can be part of a predictive or generative AI solution or work across systems to achieve the desired outcome.

¹State of AI in Telecommunications: 2024 Trends - NVIDIA Survey Report (<https://resources.nvidia.com/en-us-ai-in-telco/state-of-ai-in-telco-2024-report>)

²Telecommunications Generative AI Study September 2023- Altman Solon (https://pages.awscloud.com/rs/112-TZM-766/images/Altman%20Solon_AWS_Telecoms%20Generative%20AI%20Study.pdf)

Sample agent functions include:

- **Task Execution:** data entry, file manipulation, sending notifications, etc., based on predefined rules and instructions.
- **Event Monitoring:** monitor specified triggers or events, such as the arrival of new data or the completion of a task, and initiate the next steps in the workflow automatically.
- **Systems Communications:** interact with different systems, databases, or applications to retrieve or update information as the workflow requires.
- **Exception Handling:** handle exceptions or errors during the automation process, ensuring that workflows continue running smoothly.
- **Insights and Optimization:** Agents can collect data and provide insights into workflow performance, helping organizations identify bottlenecks and optimize processes.

Natural Language Processing

Natural Language Processing (NLP) significantly simplifies the user experience for networking by enabling intent-based communication between network administrators and the network itself. Before GenAI and LLMs, networking vendors leveraged more traditional AI/ML approaches like Naive Bayes, Support Vector Machines (SVM), Random Forests, and Decision Trees to assist in categorizing human commands and logging outputs from network systems. With the arrival of GenAI, we expect vendors to leverage transformer-based architectures and LLMs for NLP tasks due to increased convenience (no need for training) and accuracy.

NLPs play an essential role in HCI and help with the following tasks in AIOps:

- **Intent Recognition:** NLP systems can understand the intent behind human commands or requests and determine what action needs to be taken on the network, such as provisioning new resources or troubleshooting connectivity issues.
- **Intuitive Policy Definition:** Using natural language to specify access control rules, quality of service requirements, or other policies, which are then translated into machine-readable formats for implementation.
- **Automated Troubleshooting:** NLP assists in troubleshooting network issues by analyzing descriptions of problems provided by users or automated monitoring systems. NLP systems can identify critical information in these descriptions and suggest or perform actions to resolve the issues.

Intermediate Representations – Digital Twins and Network Models

Before we wrap this section, two topics crop up in many AI networking discussions: digital twins and network model schemas. Both are forms of representation of a network element (physical or virtual). A digital twin captures the state and salient operating parameters of a network element (e.g., switch, router, access point), allowing a network automation or orchestration system to rapidly view the topology of an entire network and visualize the state without the necessity to query many network objects in real-time. Digital twins can also be used for modeling and simulation – running what-if scenarios and aiding in capacity planning or predictive failure modeling. AI systems could ostensibly simulate changes on a digital twin of a network to determine potential issues before pushing out a configuration change.

Another associated concept is the network model, with a schema that captures the parameters necessary to configure a network element—one common model is the OpenConfig standard. Unfortunately, vendor support for OpenConfig is limited and not sufficiently universal. Nevertheless, a standard model can simplify configuration verification as part of an automated network configuration workflow.

While the industry works towards a standard model (which may never be achieved), AI-assisted systems today are either vendor-specific, in which they use their vendor-specific APIs to configure and query their equipment or use an intermediary network automation layer between the AI control layer and the underlying network elements. Solutions for this intermediary layer include Iential (sponsor; please check out sponsor section), IBM CloudPak for Network Automation (sponsor; additional viewpoints in the sponsor section), Gluware, and Netbrains, or custom programming with automation frameworks from vendors like Red Hat Ansible.

AI Capabilities for Network Operations

AI capabilities can be classified into a few core sets of tasks. Understanding these tasks and their application to various networking use cases helps with understanding the fit of AI to specific operations. Likewise, understanding how AI is applied provides insight into when networking vendors may offer related capabilities.

AI USE CASE MAP CORE CAPABILITIES

	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 (initially)	AI Agents	Control Modules, Copilot (initially)
Implementation	RAG	Fine-tuned FMs	Fine-tuned FMs + Multi-Agent workflow	Predictive AI, GenAI with Agents, Digital Twins
Effort	Low	Med-High	Med-High	Med-High
Timeframe (production-use)	12 months	24+ months	18+ months	Today and Evolving



avidthink.com

Discovery

GenAI and AI can offer personalized, contextualized, natural language responses to customer queries on a chat interface. This capability can apply to every step in the network automation process where a human needs to interface with the network to provide a superior user experience.

While many vendors have a human language interface today, the interfaces may need upgrading and retooling using LLMs, especially if they were designed 18+ months ago.

With retrieval augmented generation (RAG) approaches, it is relatively easy to introduce this capability by coupling vendor content to LLMs – using various search techniques (vector embeddings or other semantic search) and inserting snippet results into prompts. We anticipate this will be a critical use case for GenAI over the next 12 months. A few design principles to keep in mind with RAG:

- RAG implementations need to follow role-based access control (RBAC) principles on access
- While the approach is suitable for static data, such as documentation and configurations, it is not ideal for dynamically changing data, such as network operational status
- Data anonymization capabilities may be needed in the RAG pipeline

Design and Configuration Generation

The generative use cases where GenAI generates a design, configuration, policy, or documentation must undergo a maturity phase of extensive fine-tuning of the models before they are ready for market deployment. We expect this process to take 2-3 years to achieve maturity, depending on the complexity of the problem at hand.

The more immediate use cases for generation in the next 12 months should involve AI generating personalized marketing content or vendors using AI-copilot to assist in code generation to reduce time to market or lower the cost of product development.

Automation

Automation activities involve creating multiple tasks and workflows that require the use of agents. Adoption for discrete use cases aided by agents should begin in the next 12 months. End-to-end automation will naturally take longer to mature until each discrete task is automated. Trust in automation, where a user trusts in AI, is another barrier to adoption. It is expected that the first 12-18 months of any new automation initiative will be in the co-pilot state where the human approves and machines execute. The co-pilot stage only transitions to the auto-pilot stage after confidence is built.

Prediction

Businesses generate massive amounts of data daily, including network and security logs containing vital information about network health, user behavior, and anomaly detection. AI can parse historical data to identify opportunities for predictive maintenance and visualize findings for more accessible review. Predictive analytics solutions, as part of AIOps, use this today.

Simulation

Digital twins offer a transformative approach by creating a virtual model of a network's infrastructure, operations, and services. This digital replication allows network administrators to simulate, predict, and manage network behaviors effectively under various conditions without impacting the actual network.

AI Impact Across the Network Lifecycle

AI adoption in networking is targeted at supporting a wide range of use cases that drive simplicity and ease of use, reduce human error, cut operational costs, enhance the user experience, accelerate time to market, simulate and optimize the environment, and drive better security. This section will examine AI use in networking across deployment stages. We'll start at day 0 design activities and then proceed through day 1 deployment, day 2 management and operations, optimization, troubleshooting, and wrap up with continuous improvements. For each stage, we will describe in-market capabilities, assess the maturity of each stage, and discuss caveats and limitations that need to be addressed.

1. AI-Assisted Network Design (Day 0)

This stage covers the use of AI in automating various activities in network design, including product and service selection to meet business performance and cost objectives, including:

- **Site Planning:** AI automates site planning elements such as RF planning for optimal wireless AP deployments and power planning for data centers.
- **Topology Creation:** Automated topology generation and configuration templates from design criteria such as performance, scalability, security, resiliency, or compliance. The process is iterative and could use a natural language interface to share the user intent.
- **Automated Device / Service Selection:** AI generates optimal hardware, software, and service options from vendor offerings for performance, scale, and feature requirements.
- **Design Simulation:** AI generates synthetic data to test against performance objectives. Digital twins allow network design simulation in a risk-free virtual environment.
- **Design Optimization:** against cost, performance, resiliency, policy, and compliance criteria.

Benefits:

- **Speed:** Shorten time to deploy pre-validated or well-understood configurations.
- **Cost Savings:** Optimized product/service selection for price and performance.
- **Risk Reduction:** Reduce the risk of deployment in production by validating simulations in the lab.

Maturity of deployments: Level 0/1 (Manual/Assisted Management)

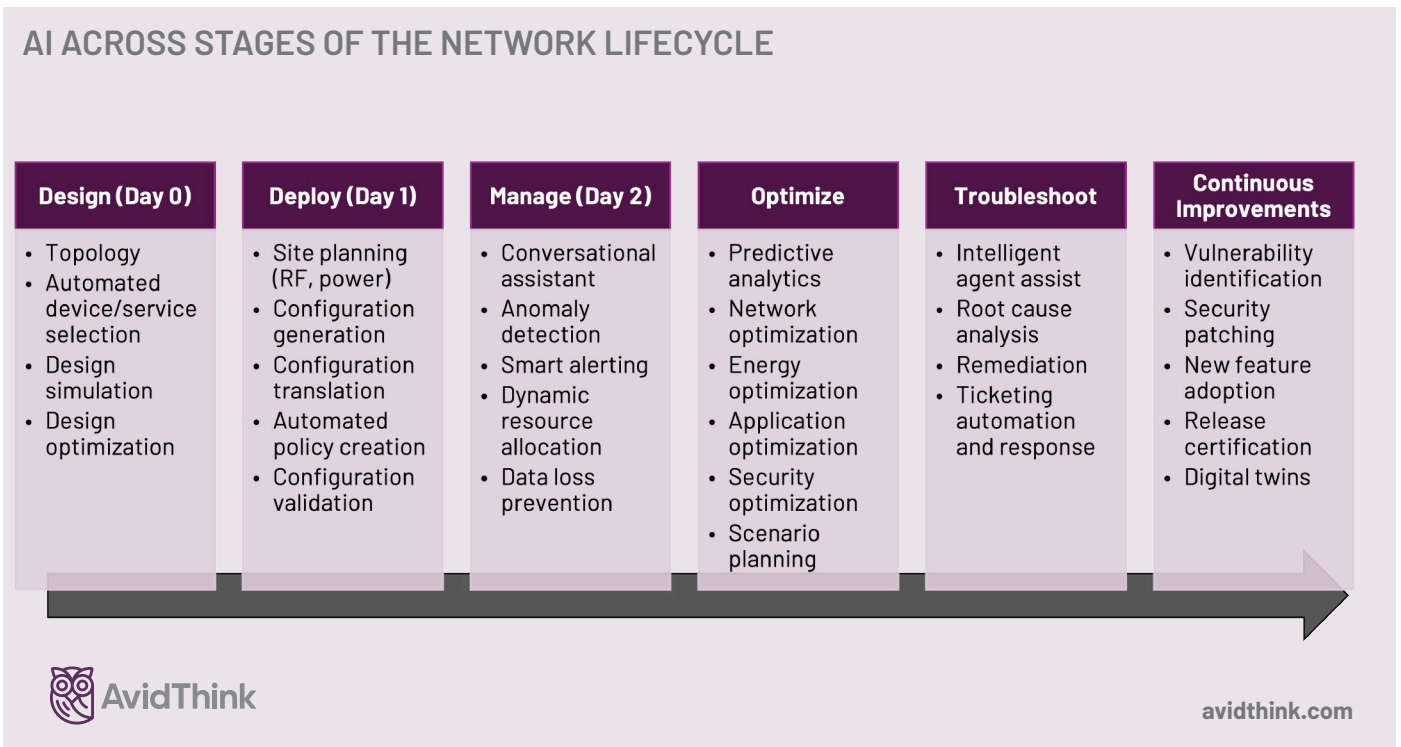
While the full vision of autonomous design would take time, subsets of the design, such as device/service selection and running multiple simulations against a model, are likely to be available through AI approaches in the next 12-24 months. We anticipate vendor-specific solutions emerging first, with multi-vendor solutions taking longer. AI-assisted network design tools for enterprise tools are in use with limited success (manual intervention is needed). Proof of concept demonstrations from significant networking vendors for other domains are available today.

Limitations and Caveats: Crafting an automated design requires integrating several systems and approaches. For instance, the AI system must aggregate data across specified intent or input parameters, existing management systems, traffic flows, and constraints to create an initial model. Subsequently, this model will have to be optimized using various AI techniques and then pass security and compliance testing.

2. Automated Deployments (Day 1)

We can break down the use cases for streamlining and automating Day 1 deployments into:

- **Configuration Generation:** AI generates the initial configuration and setup of network devices, ensuring consistency and adherence to best practices.
- **Configuration Translation:** Automate the translation of an existing configuration/API to an equivalent configuration from another vendor. If configuration changes or APIs occur across releases, the translation can also apply to a single vendor product.
- **Configuration Validation:** AI validates configurations against acceptable syntax, parameters, and adherence to compliance policies.



- **Automated Policy Generation:** AI automatically creates quality of service (QoS) or security access control list (ACL) rules through intent translation – vendor-specific policy generation from a generic intent. In addition, automated policy creation via machine learning from traffic flows – analyzing traffic flows across endpoint groups and using the input to automatically define policies to permit or deny interactions between different groups of devices, users, and applications.

Benefits:

- **Speed:** Accelerate time to deployment.
- **Reliability:** Reduce deployment errors.
- **Savings:** Address the skill-set gap and the labor shortage of skilled workers.
- **Agility:** Easily adopt heterogeneous multi-vendor solutions or easily switch across vendors.

Maturity of deployments: Level 1 (Assisted Management)

While fully automated configuration generation will take time to mature, AI can reliably generate vendor-specific network configurations using available vendor documentation and examples. Several high-value use cases in this area are feasible to implement in the next 2-3 years.

Limitations and Caveats:

Most solutions will initially be vendor-specific, which may limit their usefulness since many carrier and enterprise networks are multi-vendor. However, solutions like those from Itential that help provide an abstraction and automation mechanism between the higher-level intent and the APIs of the network elements can accelerate deployment.

GenAI will bring additional capabilities, including a conversational interface and the ability to learn over time. We’ve seen early vendor examples of using GenAI to auto-create personalized and custom dashboards specific to roles, tasks, and individual preferences.

3. Management and Monitoring (Day 2)

This stage is the farthest along in maturity. Several networking vendors have had AI/ML capabilities in their network management platforms for some time; artificial intelligence for IT operations (AIOps) has been a central use case for AI adoption in the networking domain for the last 5+ years. The term AIOps has been used to refer to network management and monitoring. However, we see the scope of AIOps in networking expanding to include other stages in the network lifecycle.

AIOps enhances Day 2 network operational support by correlating multiple data inputs, identifying problems faster, yielding quicker resolution, and, where applicable, spotting issues proactively before a user is aware. GenAI will bring additional capabilities, including a conversational interface and the ability to learn over time. We’ve seen early vendor examples of using GenAI to auto-create personalized and custom dashboards specific to roles, tasks, and individual preferences. It can enhance user experience with specific outputs such as text, audio, video, or graphics.

Combined with generative AI, these use cases are fast evolving to cover more comprehensive scenarios:

- **Anomaly Detection:** AI analyzes traffic patterns and network behavior in real time to identify anomalies and predict network failures or potential security breaches.
- **Smart Alerting:** Increased volume, velocity, and diverse telemetry sources have made it complex for a human operator to manually analyze and respond to the data to meet operational expectations. AI helps consolidate and analyze data from multiple sources.
- **Dynamic Resource Allocation:** AI-driven tools optimize network performance by dynamically allocating resources in real time based on usage patterns, ensuring optimal network efficiency and responsiveness.
- **Data Loss Prevention:** AI enhances security by monitoring data flows and user behavior, identifying potential data breaches, and preventing unauthorized access or exfiltration.

- **Conversational Assistant:** Administrators use natural language on the Network Management platform to understand the state of the network, such as the top applications or top 10 problems spotted in the past month.

Benefits:

- **Quality of Experience (QoE):** Improve end-user experience by responding to dynamic changes in real-time.
- **Cost Savings:** Save on the cost and provide improved quality of operations.
- **Reliability:** Predict network failures before they occur, allowing for preemptive repairs that minimize downtime.
- **Efficacy:** Address data fatigue by summarizing the top issues that need attention.
- **Security:** Ensures sensitive information remains protected and compliance requirements are met.

Maturity of deployments: L2/L3 (Partially Autonomous/Conditional Autonomous)

Conversational virtual assistants will be prevalent across every management platform within the next 12 months. Leading network security vendors have demonstrated these capabilities in their products, and networking vendors are likewise adopting AI and GenAI for cybersecurity and resilience use cases. Improvements with smart alerting also remain a significant opportunity for vendors over the next 18 months.

Limitations and Caveats:

AI has been used for anomaly detection for some time now. Closing the feedback loop for dynamic resource allocation, however, needs to build confidence with users before it gets adopted. In addition, turning over controls to AI to help highlight or predict what's important in the network requires a way to measure efficacy and figure out system blind spots—an open issue.

4. Optimized Environments

By leveraging AI to optimize their networks, organizations can achieve higher resource utilization, improved reliability, and better user experience, leading to a more robust and responsive network infrastructure. While solutions like dynamic resource allocation can address resource utilization in real time, AI-aided optimizations allow for future network planning and optimizations.

AI-driven network optimization can allow network administrators to optimize for multiple performance metrics and objectives simultaneously. Instead of focusing solely on maximizing throughput or minimizing latency, a broader range of competing goals, such as cost or energy efficiency, offers a holistic and adaptive network optimization approach.

Instead of focusing solely on maximizing throughput or minimizing latency, a broader range of competing goals, such as cost or energy efficiency, offers a holistic and adaptive network optimization approach.

- **Predictive Analytics:** This is the most prevalent use case for AI in networking, including planning, proactive alerting, anomaly detection for security and operations, optimization, and troubleshooting.
- **Network Optimization:** This includes capacity planning, optimization of forwarding or QoS policies, and integration with third-party systems to consider transport costs or cloud charges.
- **Security Optimization:** Analyze vast amounts of data, identify patterns, and make informed decisions to enhance the overall security posture.
- **Application Optimization:** Adopt an integrated approach that considers both network and application performance to drive superior overall application performance and ensure a seamless user experience.
- **Energy Optimization:** Conserve and optimize energy use in network environments by switching off unused ports, turning devices into a low-power mode when not in use, and providing power usage analysis across the deployment to identify power-inefficient devices.

- **Scenario Planning:** Synthetic data and the use of digital twins helps simulate various network scenarios, including potential security threats, without compromising real user data.

Benefits:

- **Cost Savings:** Optimize network resources and planning, save energy costs.
- **QoE:** Improve user experience, security, and reliability.

Maturity of deployments: Level 2 (Partially Autonomous)

Limitations and Caveats:

AI-driven optimization techniques often involve complex computational processes and require significant computational resources to train and deploy.

AI-driven optimization techniques may increase an organization's attack surface, making it more vulnerable to cyber threats and malicious attacks. Therefore, organizations must implement robust data protection measures, encryption protocols, and access controls to safeguard sensitive information and mitigate security risks associated with AI-driven optimization.

5. Troubleshooting and Customer Support

AI plays a significant role in improving customer service by enhancing efficiency, personalization, and responsiveness.

- **Intelligent Agent Assist:** Allow employees and customers to resolve issues faster with an AI-powered conversational chatbot.
- **Root Cause Analysis:** AI reduces the time needed to troubleshoot availability and performance issues by collecting and analyzing logs and event data across multiple datasets and correlating them to the baseline.
- **Remediation:** Recommend and take corrective action based on root cause analysis.
- **Ticketing Automation:** Automation of ticket generation, priority assignment and routing

Benefits:

- **Speed:** Faster time to issues resolution
- **Reliability:** Higher network availability

Maturity of deployments: Level 2 (Partially Autonomous)

Intelligent agents assist, and ticketing automation solutions are in place today and should see adoption in the next 12 months. For example, many vendors have ITSM integration with ServiceNow for automated ticketing. Improved customer service remains a high focus, and it's common to find vendors offering intelligent agents (Arista, Cisco, HPE Aruba, Huawei, Juniper Networks, Nokia, Ericsson/Cradlepoint) for assisted troubleshooting, leveraging their knowledge base.

Limitations and Caveats:

Events or telemetry data must come from a portfolio of tools for monitoring network infrastructure components and application performance. Not being able to assimilate the full scope of the environment, including network and data resident in external cloud instances, could be a gap in visibility.

Sharing support ticket insights across a global customer base requires data anonymization.

While AI can make recommendations for remediation and vendors are beginning to offer closed-loop automation, adoption today requires a human to press the trigger until trust is fully developed in AI.

6. Continuous Improvements

Networks are dynamic environments that need periodic upgrades and optimizations. Ensuring that the networks are operating with newer releases that offer the latest security patches and feature functionality remains a sore point in vendor-customer relationships that AI can help solve, including:

- **Identifying Vulnerabilities and Deploying Patches:** Machine learning can quickly analyze thousands of network devices or endpoints to validate that they have the latest software code. It can then look for potential vulnerabilities by parsing device configuration, flagging the need for security patches, and automating the patch update.
- **Identifying New Features:** Analyze the deployed code for features used and flag new capabilities as they are available in subsequent vendor releases.
- **Release Certification:** A common pain point for vendors and customers is speeding up the adoption of new code and capabilities adoption. A code upgrade for network devices in production is a tedious problem, and large corporations can take between 6-12 months of extensive testing before production deployment. With AI automating the testing of the new image against deployed infrastructure and traffic patterns, the time frame for testing and confidence for deployment can improve.
- **Testing and Simulation:** Digital twins allow for the simulation of network changes, troubleshooting, and optimization strategies in a risk-free virtual environment. Using simulation can test the impact of new deployments or changes in network configuration before applying them to the production network.
- **Security Enhancements:** Digital twins can model potential security threats and simulate attacks, allowing network security teams to observe potential impacts and evaluate response strategies. This proactive approach enhances the security posture by allowing refinement of defenses before a breach occurs.

Benefits:

- **Speed:** Fasten the adoption of the latest s/w releases and bring new capabilities into the environment.
- **Security:** Improve the security posture of the network.
- **Reliability:** Simulate improvements safely in a non-production setting.

Maturity of deployments: Level 1/2 (Assisted Management, Partially Autonomous)

Limitations and Caveats:

Machine learning to identify devices that need to be patched has been available through network management systems. Automating patch updates requires a certain level of trust in the validation system and the device recovery mechanism in case of a botched update.

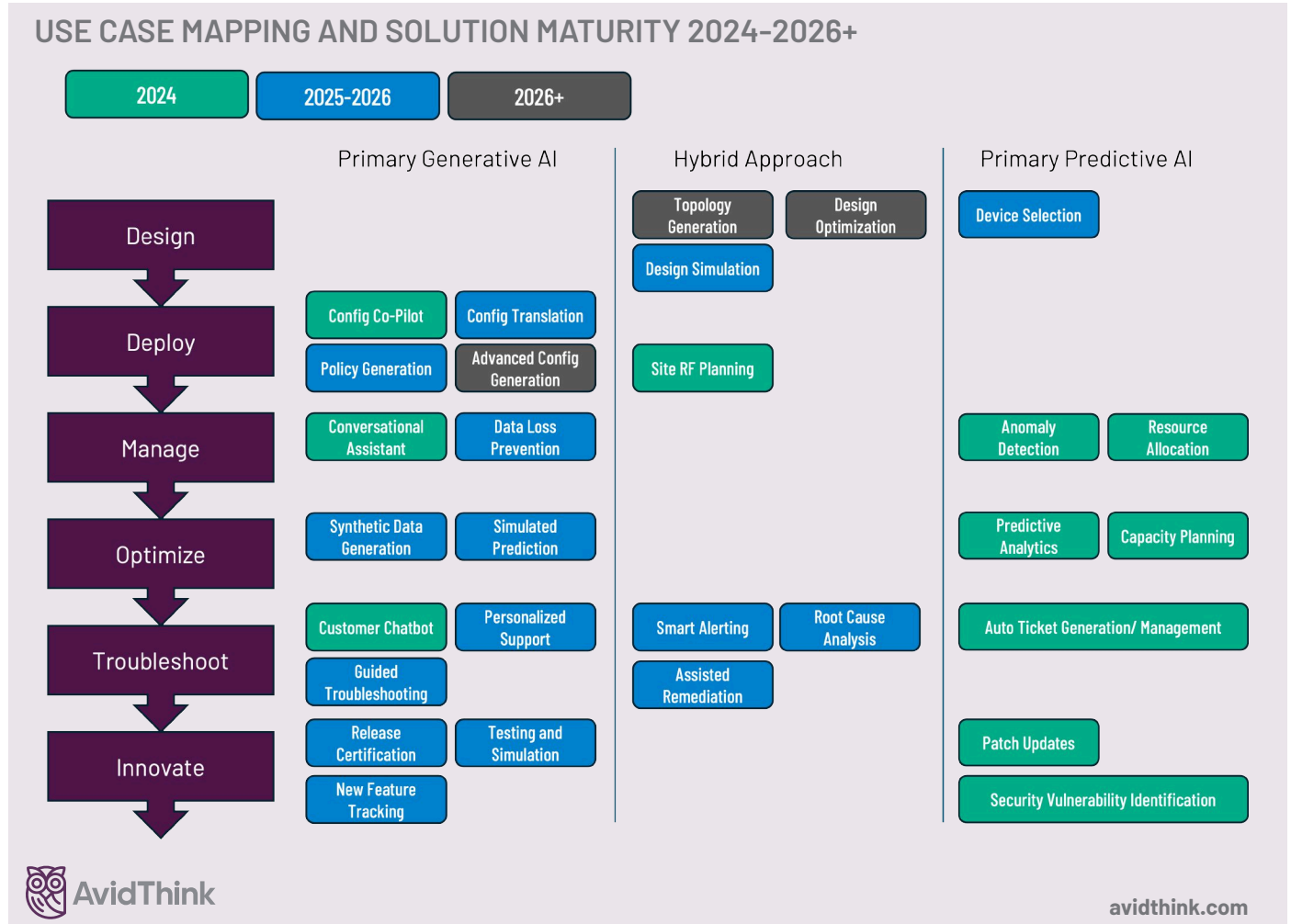
Automating QA and certification efforts to validate new code on existing deployment is an immense opportunity for vendors and customers. Vendors can successfully retire old code, freeing up resources for new development.

Digital twins allow testing network and network security change impacts in a simulated non-production environment to address issues before they are deployed in a production environment. Digital twins can be part of a continuous integration/continuous delivery (CI/CD) pipeline that verifies that the network works as expected or for "what if" analysis on modeling.

Digital twins allow for the simulation of network changes, troubleshooting, and optimization strategies in a risk-free virtual environment. Using simulation can test the impact of new deployments or changes in network configuration before applying them to the production network.

Use Case Mapping to Expected Solution Maturity

Combining the insights from the two previous sections, we can map AI’s fit to a set of network operations or activities across the different stages of the networking lifecycle. We’ve also added our best estimates as to the timeframe when these capabilities/use cases will emerge (early majority). As the market evolves, we’ll revise our estimates in future reports.



Regardless of our estimates in the diagram above, when considering GenAI’s new opportunities alongside existing predictive AI-driven solutions, readers should evaluate the value GenAI provides to their organizations relative to specific vendor strengths, weaknesses, and risks in prioritizing adoption.

Real-World Use Cases of AI in Networking

As indicated, traditional AI/ML has been used in the mobile RAN for spectrum and energy optimization and enterprise Wi-Fi for better radio resource management. Predictive ML models have also been used for anomaly detection, predictive maintenance, and capacity management. Today, with further improvements in ML/DL technologies and with GenAI, we’ve seen promising results in other use cases. We’ve picked a small subset here to showcase the diversity of areas AI/GenAI can be used:

- eBay deployed an AIOps solution for its data center network to help with incident management by focusing on predictive maintenance³. They successfully used AI to sort through all the events and logs from their network to extract predictions of potential failures and mitigate downstream failures proactively. Based on their experiences, they have sufficient confidence to allow AI to automatically generate preventative trouble tickets to improve the reliability of their network.

³“Evolving Network Operations Through the Power of ML” eBay session at ONUG Spring 2022. (<https://onug.net/events/evolving-network-operations-through-the-power-of-ml/>)

- Nokia (report sponsor, check out their perspectives in the sponsor section) helped telecom operators like stc achieve energy savings on their RAN, and likewise with KDDI. KDDI used Nokia's AI-powered AVA (Automation, Visualization, Analytics) platform for Energy Efficiency to analyze and anticipate changing traffic volumes in the sites and cells of its RAN, allowing powering down of cells to reduce energy use. On average, KDDI reduced power consumption by up to 50 percent in low-traffic environments and up to 20 percent per cell.
- Cox used GenAI (Anthropic on AWS) to pilot Service Health, its analytic models that predict and prevent network issues. The model uses unstructured data, including anonymized customer transcripts from online chats, calls, and tech journal entries. Through Service Health, Cox reduced problem resolution times from 45 minutes to 45 seconds in many cases.
- China Mobile, Zhejiang, used Huawei's ICN Master solution with AI to auto remediate and close out 80% of their 1000+ trouble tickets/month with over 90% accuracy in the system's ability to figure out the root cause.
- Google is leveraging AI to enhance its network operations, particularly in incident analysis and resolution. It has developed an AI-powered assistant to help its internal teams analyze network incidents more efficiently across its global network, which spans 32-38 regions and over 200 countries. Its AI assistant, which utilizes LLMs and search technology, analyzes postmortem documents from previous outages, providing engineers with summaries, root cause analyses, and impact assessments without needing to sift through extensive logs and documentation manually.
- Motor City Wash Works plans to use Cradlepoint's (an Ericsson subsidiary) NetCloud AIOps Dashboard and AI-based NetCloud Assistant (ANA) to manage their network of industrial devices installed in car wash locations nationwide. ANA uses NLP to provide ad-hoc queries on network operations, and NetCloud AI utilizes ML to help with fault identification and root cause analysis.

Networking Vendor AI Implementations

Networking vendors that serve enterprises and carriers have all introduced AI/ML and, more recently, GenAI capabilities into their offerings. Many GenAI enhancements today are flagged as beta or proofs-of-concept—we anticipate that once vendors can improve the reliability of the features in the next 12-18 months, they will be productized.

Leading networking firms in the enterprise and data center market include Cisco, HPE Aruba, Juniper (in the process of merging into HPE), Arista Networks, Extreme Networks, Huawei (outside the US and EMEA markets), and VMware (virtualized and edge networking). All vendors have announced or are shipping AI capabilities – whether traditional AI or GenAI. For this section, we will pick a few examples to cover. Readers looking for a more comprehensive and detailed analysis should contact us directly.

We start with **Juniper Networks**, considered by many enterprises to be a leader in AI-driven networking because of their Mist AI platform, which comes from their acquisition of Mist Systems in 2019. Juniper Marvis, an AI-powered virtual network assistant that can detect anomalies, describe network problems, perform root cause analysis, and assist in troubleshooting, has powered Juniper's Wi-Fi and campus offerings. Marvis virtual network assistant (VNA) has been extended to SD-WAN and routing domains and for the data center (integrates with Juniper's intent-based Apstra solution).

HPE's messaging has bolstered the perceived leadership of Juniper in AI for networking; HPE has shared that a critical reason for its acquisition of Juniper is Juniper's Marvis AI capabilities. While HPE **Aruba** already had AirMatch, an AI-based RF optimization solution, and AI analytics, HPE's AI capabilities were likely not on par with Juniper's. Nevertheless, HPE Aruba is also working on integrating GenAI LLMs into their HPE Aruba Networking Central platform. HPE hopes to improve network performance, accuracy, and security by providing real-time, context-specific responses to user queries.

Meanwhile, **Cisco** is the market share leader in enterprise networking, messages AI capabilities, and vision at their recent Cisco Live! US event. On the AIOps front, Cisco described new Cisco ThousandEyes capabilities and AI-native workflows in Cisco Networking Cloud to deliver Cisco Digital Experience Assurance. Cisco likewise indicated it would use recently acquired Splunk's analytics capabilities to power its AI-enabled network operations. The improvements in observability coupled with AI intelligence are anticipated to provide improved insights and allow for faster root cause analysis and remediation of end-to-end traffic across both networking assets owned by the company and those traditionally outside their span of control (e.g., carrier and cloud networks).

Other enterprise networking vendors include **Arista**, which has developed features in its EOS networking software to provide advanced analytics. It also uses AI/ML in its Wi-Fi solution radio resource management. In addition, Arista EOS aims to use its Arista Autonomous Virtual Assist (also called AVA, like Nokia's) capability coupled with their multi-modal data lake that exposes real-time network telemetry (from Arista and partner network elements), to provide insights and aid in human decision making. Another vendor is **Extreme Networks**, which recently showed off their ExtremeCloud IQ CoPilot, a network assistant with AI and digital twin capabilities equipped with explainable AI to engender trust with its users on how it arrived at a decision.

In addition to the enterprise networking companies above, leading IT solutions provider **IBM** is looking to apply expertise from its IBM watsonx AI platform to its network orchestration and automation solutions. These solutions include CloudPak for Network Automation and its recent acquisition, Pliant, an IT automation and orchestration solution. IBM also owns Red Hat with its Ansible automation platform, which has a strong presence in the networking space. Plus, IBM is currently in the process of acquiring HashiCorp, a leader in infrastructure and security lifecycle management for hybrid clouds, which will provide it with greater access to infrastructure automation systems. On the other end of the company size spectrum, we have network AIOps startups like **Augtera**, who have demonstrated success with Fortune 500 eCommerce customers on the enterprise front and Orange in the carrier space. Similarly, we have startup **Aviz Networks**, which leverages the open-source network operating system, SONiC, and has launched a GenAI-powered networking copilot solution.

Integrating AI in networking is an ongoing process, with all networking companies racing to understand better how to improve their traditional predictive AI capabilities while leveraging GenAI in their products. We expect the rest of 2024 and 2025 to see even more AI-focused announcements as vendors seek to better each other.

As we transition to look at carrier solutions, we note that AI/ML use will become increasingly pervasive across all product portfolios for networking vendors. For example, **VMware** utilizes AI for network security with their data center networking virtualization product line, VMware NSXi, and offering assisted network operations with VMware Edge Intelligence – an AIOps-based network analytics and IoT security platform. VMware also provides a platform for bringing AI/ML into mobile networks with VMware RIC (RAN Intelligent Controller), an Open RAN platform that provides RAN intelligence and programmability. VMware RIC hosts near-real-time applications (xApps) and non-real-time applications (rApps) that bring energy efficiency, network optimization, spectrum optimization, security monitoring, and other capabilities to the mobile RAN. VMware encourages partners to build apps on top of their RIC: partners **Aira Technologies** provides AI-powered spectrum and energy optimization, **Rimedo Labs** enables traffic steering, and **NetAI** enables traffic forecasting and resource autoscaling.

Earlier, we touched on **Nokia's** AVA AI-powered solution for network optimization and management. Nokia continues to innovate in AI-driven networking across multiple domains. They've integrated a ChatGPT-like interface into their SR Linux OS, allowing for natural language commands on hardware routers, reducing the

need for specialized CLI knowledge. Their NSP IP domain controller now includes augmented assurance capabilities, leveraging AI/ML to predict and mitigate potential disruptions in IP transport networks. Additionally, Nokia's Deepfield Defender utilizes AI/ML, internet-scale mapping, and programmable router silicon to provide automated DDoS protection, detection, and mitigation. These advancements, along with their ongoing work on Natural-Language Networks and collaboration with NVIDIA, demonstrate their ongoing commitment to AI-powered networking solutions.

In addition to the capabilities of **Cradlepoint** we cited earlier, parent company **Ericsson** has been pushing AI-driven networking with their Ericsson Operations Engine, an AI-powered platform for network operations and management, and touts their Cognitive Network Operations, which uses AI for predictive maintenance and automated troubleshooting. As with other RAN providers, Ericsson also has AI-powered RAN optimization that improves network performance and energy efficiency. More recently, Ericsson deployed their AI-powered intent-based solution at DNB, Malaysia's 5G wholesale network, that helped avoid SLA breaches through better resource utilization and AI-enabled planning.

Huawei has been heralding its AI efforts for many years as part of its SDN expansion. Huawei's Autonomous Driving Network (ADN) describes its AI-driven vision as encompassing multiple networking domains. Many carriers use their ICN and Net Master platforms for automation and intelligent operations. They also market iMaster NCE, a suite of AI-powered network management and control systems that spans campus, data center, and access and optical networks.

Integrating AI in networking is an ongoing process, with all networking companies racing to understand better how to improve their traditional predictive AI capabilities while leveraging GenAI in their products. We expect the rest of 2024 and 2025 to see even more AI-focused announcements as vendors seek to better each other.

Barriers to AI Adoption

For AI to drive a wholly autonomous network in the future, there remain a number of impediments that have to be addressed by the industry – vendors, enterprises, carriers, and consultants and integrators:



- **Trust in technology:** Generative AI applications suffer the risk of hallucination, and concerns about proprietary data ending up in the public domain exist. Matching the right AI approach to the proper use cases and securing data during implementation will help build trust.



- **Scalability of RAG and today's workflows:** While a number of GenAI use cases using RAG have been tested in demo environments, their ability to scale in production remains an unknown. For instance, context length available from different LLMs constrains the applicability for some use cases. More recently, AI experts have questioned if the RAG architecture is the best solution for incorporating proprietary or domain-specific data in GenAI workflows, especially when accuracy is important.



- **Explainability of AI:** Transparency in an AI algorithm's decision-making capabilities is crucial to trust and the evolution of its role from an advising co-pilot to a trusted automation system. Vendors must develop interpretable AI models and algorithms to explain their decisions and recommendations transparently and efficiently.



- **Periodic model retraining:** While AI implementations can improve their performance from fine-tuning, fine-tuned models may need periodic retraining. This adds to the solution's cost and the response's efficacy.



- **Fragmented networks:** AI deployments for end-to-end service management and orchestration must support multiple vendors and administrative domains, including the business, service, and resource layers. This could delay the full implementation of autonomous deployments. Still, solutions that provide cross-vendor orchestration and automation from vendors like Iptential or Ansible from Red Hat/IBM could mitigate this constraint.



- **Data privacy:** As we have seen with cloud deployments, equivalent data privacy and data residency rules will be needed for AI use. This is particularly important for AI solutions that use external APIs to access FM services.



- **Holistic replanning:** Selective use cases focusing on a single part of a more extensive legacy process may show good near-term results but need revamping to achieve fully autonomous implementations. Companies might see greater value if they can re-engineer entire end-to-end processes and workflow versus having piecemeal, siloed, deployments.



- **Human friction:** Incorporating AI into legacy processes built around the needs and capabilities of human workers can lead to disjointed rollouts and friction for employees.



- **Unclear ROI:** There's less doubt about the return on investment in predictive AI/ML technologies—they cost less to train and have shown value over the past decade. However, GenAI is immature and prone to hype, leading many companies that were initially bullish to take a more cautious path, which may slow down innovation and adoption.

Impact of AI on People and Organizations

Organizations will need to consider the longer-term societal implications of AI/ML and GenAI. Consultancies like McKinsey, BCG, and Accenture estimate that GenAI will impact a large number of job roles—for example, BCG estimated that 90% of tech jobs will be directly impacted by GenAI⁴. And the International Monetary Fund (IMF) has estimated that about 60% of jobs in advanced economies could be affected by GenAI⁵.

Given GenAI's immaturity and rapid evolution, the future of GenAI and AI/ML is unclear. What's clear is that AI/ML and GenAI have a concrete impact on network operations—long-term use of traditional AI/ML has yielded consistent benefits over time, and early pilots with GenAI are promising.

Whether and when we will achieve fully autonomous networks is uncertain. But we know the impact on organizational transformation will be profound, as traditional roles will evolve, requiring a workforce skilled in both networking and AI technologies.

Organizations adopting AI in networking must help their employees adapt to new roles, focusing more on strategic oversight and less on routine network management, fostering a culture of continuous learning and innovation.

GenAI and AI-assisted operations could also impact the industry, blurring roles and responsibilities in converging previously siloed functions such as NetOps and security operations (SecOps). AI can augment specialized skills, allowing non-specialists with solid reasoning and management skills to benefit from AI-enabled expertise.

A related concern that network veterans have suggested is the potential loss of deep expertise that comes from decades of hands-on experience in configuration and troubleshooting. Suppose a new generation of network operators becomes too reliant on AI-powered systems or as networks become autonomous. In that case, there are concerns about the lack of long-term innovation if AI systems are limited to the currently available base of knowledge built up by the previous and current generations of networking experts.

Recommendations for Enterprises and Carriers

Enterprises and carriers who take a measured and careful approach to AI in networking can navigate the complex landscape. The key is to remain flexible and adaptable as the technology evolves rapidly, balancing the potential benefits with the associated risks and challenges. Here are our recommendations:

- 1. Start with proven AI/ML solutions:** Begin by implementing well-established predictive AI and machine learning solutions, particularly in areas like anomaly detection, capacity planning, and energy optimization. These have demonstrated consistent benefits over time and can provide immediate value.
- 2. Adopt a phased approach to GenAI:** While GenAI shows promise, it's still maturing. Start with low-risk use cases like conversational interfaces for network management platforms or assisted troubleshooting. Gradually expand to more complex applications as the technology proves itself.
- 3. Focus on use cases with clear ROI:** Prioritize AI implementations that address specific pain points or offer tangible benefits, such as reducing Mean Time to Repair (MTTR), improving energy efficiency, or enhancing security posture.
- 4. Invest in data infrastructure:** Successful AI implementation relies on high-quality, comprehensive data. Invest in robust data collection, storage, and management systems across your network infrastructure.
- 5. Embrace digital twins:** Leverage digital twin technology to simulate network changes, capacity planning, and security testing risk-free. This can reduce the potential for errors in production environments.
- 6. Prioritize security and privacy:** As AI systems require access to sensitive network data, ensure robust security measures and data privacy controls are in place. This is crucial when using external AI services.

⁴R. Ebeling, A. Puget, S. Ricard, and D. Sparkes-Wallace, "Tech Leaders Need to Rethink Talent Strategy for GenAI," BCG Global, Apr. 17, 2024. <https://www.bcg.com/publications/2024/leaders-need-to-rethink-tech-talent-strategy-for-genai> (accessed Aug. 01, 2024).

⁵"AI Will Transform the Global Economy. Let's Make Sure It Benefits Humanity.," IMF, Jan. 14, 2024. <https://www.imf.org/en/Blogs/Articles/2024/01/14/ai-will-transform-the-global-economy-lets-make-sure-it-benefits-humanity> (accessed Aug. 01, 2024).

7. **Foster a culture of continuous learning:** As AI transforms networking roles, encourage your workforce to develop skills in both networking and AI technologies. This helps smooth the adoption of AI-driven systems and reduce resistance to change.
8. **Collaborate with vendors:** Work closely with your networking vendors to understand their AI roadmap and how it aligns with your needs. Participate in early adopter programs to influence product development.
9. **Consider multi-vendor strategies:** While vendor-specific AI solutions may offer deeper integration, consider multi-vendor approaches to avoid lock-in and benefit from best-of-breed solutions across different network domains. Establish intermediate mediation layers that integrate vendor-native APIs while allowing a higher-level abstraction for multi-vendor orchestration and control.
10. **Implement robust governance:** Establish clear policies and procedures for AI implementation, including guidelines for data usage, model training, and decision-making processes.
11. **Start building AI expertise:** Start cultivating in-house AI expertise to evaluate vendor solutions better, customize implementations, and potentially develop proprietary AI applications for your unique network environment.
12. **Plan for the long term:** While aiming for quick wins, develop a long-term strategy for evolving towards autonomous networking. This includes consideration for workforce transformation, technology evolution, and potential industry shifts.
13. **Maintain human oversight:** Even as AI capabilities grow, ensure human experts remain in the loop, especially for critical decisions. This helps build trust in AI systems and maintains a deep understanding of network operations.
14. **Evaluate and iterate:** Regularly assess the performance and impact of AI implementations. Be prepared to adjust your approach based on real-world results and emerging technologies.

Wrapping Up

To ensure that the industry's collective vision for autonomous networking does not remain just a pipe dream, we need to further advance and integrate the use of AI across all aspects of and domains in networking. AI has, and will, continue to transform how networks are designed, deployed, managed, optimized, and secured. AI, ML, and GenAI enhance network efficiency, reliability, and the user experience. They enable a more scalable and adaptable network infrastructure to meet the demands of large-scale 5G, IoT, modern hybrid cloud and distributed user and endpoint deployments.

However, widespread adoption of AI in autonomous networking faces several barriers the industry must address. These include technical challenges such as ensuring data security and privacy, the need for robust AI approaches that can handle diverse and dynamic network environments, and overcoming resistance to change within organizations. Addressing these barriers is crucial for realizing the full potential of AI-driven networks.

Ultimately, the vision of a fully autonomous network will be in continuous development during a 5-7-year journey that will bring early benefits along the way – AI/ML has already delivered value, and GenAI, in the form of assistive copilots can provide incremental value with low risk while we mature the technology.

We will continue tracking this space's evolution over the next 12 months and, given the rapid innovation in GenAI, look forward to revisiting this topic in 2025. In the meantime, we encourage you to reach out to us for a deeper conversation 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.



Interview with Moshe Lavi, Director of Technology Product Management

Our report covers the potential of AI/ML to optimize telco networks, particularly in areas like performance, efficiency, and sustainability? From VMware by Broadcom's perspective, could you provide some examples, perhaps in the context of 5G radio networks?

Communication service providers (CSPs) are struggling - costs are skyrocketing and ARPU has flatlined. They're looking to AI and ML to reduce their operational expenses and improve their cost structure, while play a larger role in revenue generation down the road.

How?

While AI is going to help CSPs do more with less, in general - supporting everything from sustainability improvements to extending the lifespan of network equipment and optimizing resources, let's dig into a few examples.

Today cell sites are deployed on a virtual stack, creating layers of abstraction. In these virtualized environments, each network is made up of of four different abstracted layers. With this level of granularity, the user can expect a variety of metrics and alarms to arise. With AI onboard, VMware solutions help CSP better analyze and get a grasp of the metrics. ML learns from faults to discover trends and identify root cause.

We then take it a step further to involve AI to suggest ways to fix a failure. In many cases, CSPs aren't sure what the right resource is to solve problems they see, but AI can provide insight that their teams don't possess yet.

On the automation side, onboarding artifacts into automation is often a significant expense. The users and the vendors need to spend hundreds of hours on configuration. With VMware, we leverage AI and ML functionality to configure artifacts dynamically into their environment and then support ongoing operation. This benefit extends throughout a deployment, saving hundreds of hours of work from adapting artifacts for different locations, different networks, and different hardware.

Our solutions pre-analyze the environment to avoid cases where an upgrade or day two operation, like scaling, will fail. In some cases, we see pre analyzing and pre verification of the environment will help the user save time by avoiding failure troubleshooting.

These tools will change CSP cost structures to allow them to realign resources, and then focus on generating new revenue, instead of handling operations.

Looking towards the future of networking services, how is AI facilitating the development and management of next-generation offerings? Are there any specific emerging technologies or services where AI is playing a crucial role?

The role of AI as networks become more autonomous and self-healing will continue to advance, critically enabling next generation services. In addition to enabling several use cases the industry has been working on for some time now, like network slicing and automated site planning, CSPs should also expect to have more intelligent workload placement and distributed infrastructure management through edge computing.

In fact, a significant amount of the growing impact of AI in CSP networks will be felt in operations. They will be able to detect anomalies faster and more accurately than traditional architectures. Automated deployment of sites and workloads at scale will be the norm as will steering traffic based on real-time intelligence of an end-to-end network. Networks will be self-optimizing, identifying patterns from unified data to detect and predict network anomalies for proactive issue resolution. Unified management across the network will provide a single pane of glass for onboarding and deploying network functions to lifecycle management of NFs and AI models.

AI will also play a role enabling other emerging technologies. We'll see enhanced management and provisioning of technologies like eSIM management, digital-twins for simulation and optimization. All of this while also leveraging network data to continue to reduce carbon emissions and consume less energy.

Most importantly, the impact of AI will also be felt by telco customers. It will enable personalization of services, predict and preempt issues before they impact customers, and optimize service quality based on individual usage patterns. This will result in improved customer experience.

How do our readers learn more about VMware by Broadcom's solutions utilizing AI/ML in networking?

Visit the VMware website and reach out to your sales representative, to explore the ways we leverage (and enable our customers to leverage) AI/ML.

Q&A with Chris Wade, Co-Founder & CTO, Itenial



We've been using algorithms and automated scripts to optimize network operations for a very long time. What's changing with today's AI/ML and generative AI?

The ability for AI/ML/GenAI to impact network operations has many of the same dependencies that

robust automation does. Programmability, deployment patterns, accurate data, integration with operational systems, etc. are all required to generate and maintain network configurations. There is significant potential that GenAI could build the automations that we develop today.

There is a cohort who believes that we can maintain current engineering standards with legacy CLI interfaces and achieve AI Operations for networks. I am not one of them. While GenAI has much promise, we believe expert systems provided by network vendors will be the dominant paradigm. Programmability brought us controllers and cloud services. These software systems are adding telemetry and data capabilities, leading to expert systems with AI embedded. As this becomes standard, much like programmability, this will enable an ecosystem of AI to form. Itenial is focused on taking advantage of these expert systems to advance automation in support of more advanced operational patterns.

As networks become distributed across multiple domains and many locations, virtualized, and API-driven, their complexity has grown substantially. Do you think AI becomes a necessary tool for ensuring ongoing scalable and secure operations?

The short answer is yes, but the industry will follow a common pattern when dealing with complex systems by focusing on domains. If automation is an analog for genAI, we have seen automation thrive when constrained to a single domain (cloud, WAN, DC). There is a long list of failed NMS products that attempted agnostic abstraction that couldn't manage complex networks.

Two additional complexities you highlighted were velocity and security. Applications and users require more changes every day, in addition to maintenance activities growing rapidly to maintain a secure environment.

Automation today is a static activity. We attempt to build automation in a reusable and scalable way, but it needs to be maintained as network conditions morph and multiply. As complexity escalates, we will view the need for GenAI

much the same way as we looked at early computing being necessary to manage other infrastructure.

Is network automation a necessary condition for AI/ML to add value? And is it a sufficient condition? If not, what else is needed from the network to let AI/ML do its magic?

The first use case for ML in networking was focused on telemetry. Being a large data problem, it made sense to use ML for data forecasting and analysis. It is also useful in assurance to pattern match against historical outages and performance issues to identify root cause and dependency among network issues.

For GenAI to move into the management and control plane, there is significant domain expertise required. Maybe we call it SpecificAI (just kidding). The metaphor I think about is, who will build the most robust Cisco Smart Agent? Will it be a software company who scraped the internet to build an LLM or trained on a handful of networks? Or will it be the company who has every ticket ever opened, every network config supplied by customers, and all best practices defined by standard architectures and validated designs. As AI becomes a feature and not a product, it will be embedded in all platforms – Palo Alto, Cisco, AWS, etc. This translates into AI in networking focused on domains rather than a GenAI system that can understand all networks.

Most networks are brownfield. That's stymied and added friction to network automation. Is it the same with the state of the art in AI? Will AI and AIOps be slowed down due to brownfield deployments and legacy equipment?

We will need to make determinations on where to focus our attention. There are attributes of network domains and devices which make them more feasible to support both automation and AI concepts. Devices with modern programmability, telemetry, and possibly a controller or cloud service are high on the list. Domains such as Cloud & DC with high cycle rates or domains with modern approaches like SD-WAN are priorities as well.

If we attempt to put AI on networks rather than domains, I fear we will repeat past failures. There is no magic button that will take the undocumented, bespoke, snowflake, multi-vendor network and turn it into a lights-out, closed loop operating model.

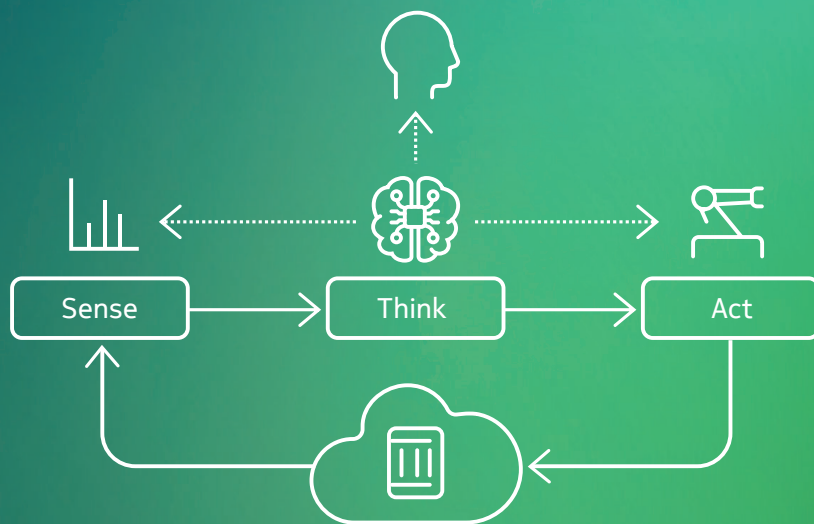
We have a technology that can transform how we operate all infrastructure. This can be the most transformational operating improvement we have ever experienced. We need to work together to build programmability into our AI strategies so that everyone can participate more than we do today via prompt engineering.

Discover the exponential potential of networks, built for AI, with AI

AI and networks rely on each other.

At Nokia, we build networks that sense, think and act:

- AI enhances analytics to make sense of data
- AI assists humans in decision making
- AI complements automation to turn insight into actions



[Click here to know more](#)

NOKIA

Interview with Cory Wepler, Head of Technology Strategy for IP Network Automation, Nokia

How does Nokia view the role of AI in networking?

We see AI playing a central role in the evolution towards autonomous networks. AI today already plays a major role in making sense of the data to generate operational insight. Moving forward, AI can also assist humans in the decision-making process, and augment automation to turn insight into action.

Why isn't AI more widely adopted at telcos today?

As we investigate, develop and deploy some of these AI capabilities with our customers, we realize that AI is not the solution to solve every problem, which can lead to a form of disillusionment that has been slowing down its adoption. However, AI and machine learning can be extremely powerful at identifying patterns, detecting outliers, making predictions and driving decisions. The success of AI relies on a comprehensive strategy

that takes into account organizational, data, tools and operational aspects to ensure alignment with the business objectives & processes.

Do you think generative AI can have a bigger impact?

We're just starting to embark on the generative AI capabilities, but I believe it will be impactful and transformative. It has the potential to redefine the human to machine interface through assistance, and significantly simplify some of the more complex tasks that we do today, like programming the network and installation and troubleshooting. Generative AI will democratize the use of the automation by reducing the barrier of entry.

Do you see any constraints around generative AI?

AI models require access to vast amounts of relevant data, but operators are cautious about exposing their sensitive data to AI models. Our commitment to data privacy and security is paramount. Also, this is a rapidly evolving technology, so we need the flexibility to change AI models to ensure we always use the most efficient solution for the problem we want to solve.



Interview with Rahul Aggarwal, CEO and Co-Founder of Augtera

What exactly does Augtera do to transform network operations with AI and machine learning?

At Augtera, we apply AI and machine learning to operational data to prevent failures in network and IT infrastructure operations, eliminate noise, and dramatically reduce the time it takes to remediate and find the root cause of failures. We're focused on bringing AI-driven insights to network operations, from data centers to large-scale LLM infrastructure, all the way to service provider networks. Our platform is software-only, scalable, and deployable on-premises or as SaaS, making it versatile and adaptable to different environments.

How does Augtera differentiate itself from open-source AI solutions?

When we started Augtera, we experimented with open-source algorithms, but they produced a lot of noise without actionable insights. That led us to develop our own purpose-built algorithms, tailored specifically for networking environments. These algorithms have much higher fidelity and actionability, allowing us to deliver significant improvements in operational efficiency and reliability at scale.

What's your perspective on the role of LLMs in network AI ops?

While LLMs have garnered significant attention, they're just one piece of a larger puzzle in network AI ops. Our core algorithms are unsupervised and don't rely on LLMs. We see LLMs playing a role in recommending remediation based on public domain data and facilitating human language query interfaces. However, the bulk of our AI capabilities, which we've been developing long before the current LLM movement, focus on pattern recognition in metrics and logs using unsupervised machine learning and natural language processing.

How does Augtera's AI handle the massive amount of data from complex infrastructures like data centers and 5G networks?

We've designed our platform to handle large volumes of data across various layers of infrastructure, from the cloud and data centers to WAN and 5G networks. We've built purpose-built algorithms after finding that open-source solutions weren't effective with production data. Our platform is software-only, can be deployed on-premise or consumed as SaaS, and scales horizontally. It's designed to handle big data and telemetry streaming, supporting various data inputs like SNMP, Syslog, open config, gRPC-GNMI, JSON logs, and Kafka. We're agentless but offer an agent for additional value on Linux systems. Augtera applies unsupervised algorithms to sift through this data in real-time, finding anomalies, predicting issues, and automatically correlating events. Our comprehensive approach allows us to correlate data across multiple layers and drive actionable insights. This reduces noise and cuts down on unnecessary tickets. For example, we've helped customers reduce tickets by 70 to 90 percent and cut down detection time from 47 minutes to just 1 minute.

Can you give examples of how Augtera's AI improves network reliability and efficiency?

Absolutely. One of our common use cases is the early detection of optical misbehavior, which is particularly crucial as optics become more prevalent in servers. We can detect these issues days in advance, allowing operators to perform maintenance before they impact applications. Another example is our ability to correlate network changes, such as a BGP change, with resulting traffic bursts and congestion. We can also detect application misbehavior using TCP retransmits and correlate it with network congestion. This helps pinpoint the root cause of issues quickly, avoiding the typical finger-pointing between teams.



Q&A with Andrew Coward, General Manager for Software Networking

Andrew, what's the current state of AI in the networking industry, and why aren't we seeing the promised results?

We're in a fascinating period of transition in the networking industry regarding AI. However, the promises of AI aren't being fully realized because, in many cases, we're taking the wrong approach. The challenge lies in the mismatch between the current focus on Large Language Models (LLMs) and the nature of network data, which is primarily time-series data and logs. These two don't naturally work well together, which explains the lack of great results we're seeing in network AI applications.

How does IBM propose to address this challenge and improve AI's effectiveness in networking?

At IBM, we believe the key is to combine different technologies. We need to integrate various AI methods and statistical approaches alongside LLMs to get the right answers. Our experience over the past 20+ years has shown that one model or method isn't sufficient. We anticipate a hierarchy of models will be necessary, each addressing different aspects of network analysis and problem-solving.

For example, we have been working on a time series model which we published as open source on the Hugging Face hub. The model, called TinyTimeMixer, has been downloaded over 500,000 times since April. Clearly, time series data models are needed not just in networking!

Can you provide an example of how this multi-model approach would work in practice?

Consider a scenario in a mobile operator's call center where a customer complains about Instagram or TikTok not working correctly. To address this, we need to identify the user, their device, and recent activity. Then, we must systematically check various network elements to isolate the issue. This process requires different models for each step.

LLMs alone aren't suitable for analyzing time-series data, so we first apply graph or other AI technologies to process this

data. We then bring these insights to an LLM to create a coherent narrative of what happened. This hierarchical approach allows us to build a complete picture of the problem, whether it's in the radio network, internet gateway, or a combination of factors.

What are the challenges in implementing this approach, particularly in telco networks?

The primary challenge is the sheer volume of data. Telco networks generate terabytes of data very quickly. We need to aggregate this data in real-time, shrink it, and make decisions rapidly at a massive scale. This requires a distributed approach to data collection and processing, with AI algorithms running at various points in the network to pre-process and understand the data before centralization.

Another significant challenge is the automation framework. Many telcos still rely heavily on manual intervention for configuration changes. For AI to truly add value, we need robust automation capabilities throughout the network, allowing AI-driven decisions to be implemented efficiently and safely.

How do you see the future of AI in networking evolving, and what role will IBM play?

We're at an exciting juncture. The future of AI in networking will involve a comprehensive approach that combines telemetry, automation, and various AI technologies. At IBM, we see ourselves at the forefront of these technologies. Our focus is on helping our customers implement the necessary infrastructure to fully leverage AI's potential in reducing operational costs, decreasing time to fix issues, and increasing customer satisfaction.

There's still much work to be done in each of these areas, but we're passionate about driving this transformation. Over the next few years, we'll be dedicated to putting the right infrastructure in place to enable AI to deliver on its promises in the networking domain.



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

©2024 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. 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.