AI operations for IP and data center networking

A journey towards autonomous networks

White paper

Service providers everywhere are embracing TM Forum's framework for autonomous networks. They want to build intelligent, self-managing communication networks that can unleash new possibilities by delivering fully automated "Zero X" experiences—meaning zero wait, zero touch and zero trouble—to all users and consumers.

Artificial intelligence (AI) promises to help service providers succeed at every step of this journey by supporting faster, smarter automation that enables their networks to sense, think and act. Read this paper to learn why AI is essential for making autonomous networks a reality and discover key AI concepts and use cases that will help service providers generate more value at every stage of the network lifecycle.

Contents

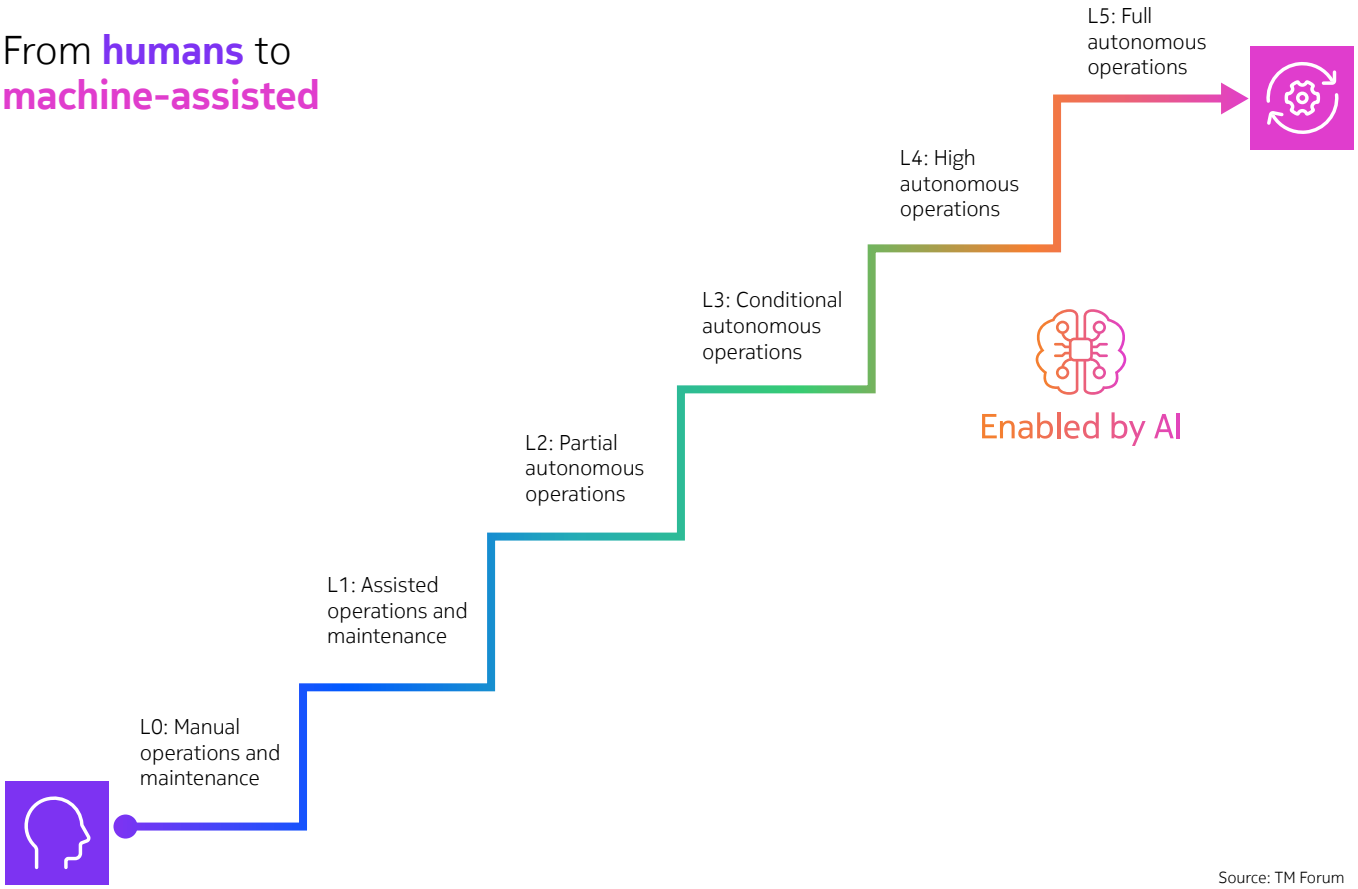
The promise of autonomous networks

Introduction

Service providers worldwide are making the journey towards fully autonomous networks as outlined in a framework being developed by TM Forum (TMF). The Autonomous Networks framework refers to intelligent, self-managing communication networks that can independently perform tasks such as configuration, monitoring, optimization and troubleshooting without human intervention. The concept builds on automation and network intelligence, aiming to create networks that are automated, but also self-driven and self-optimized.

The TMF model defines six levels of autonomy to be achieved through network evolution, and this is where artificial intelligence (AI) truly shines. AI and network automation are becoming increasingly intertwined, with their close relationship driven by the complexity of modern networks and growing demands from users.

Figure 1: TMF Autonomous Networks framework – Six levels of autonomy



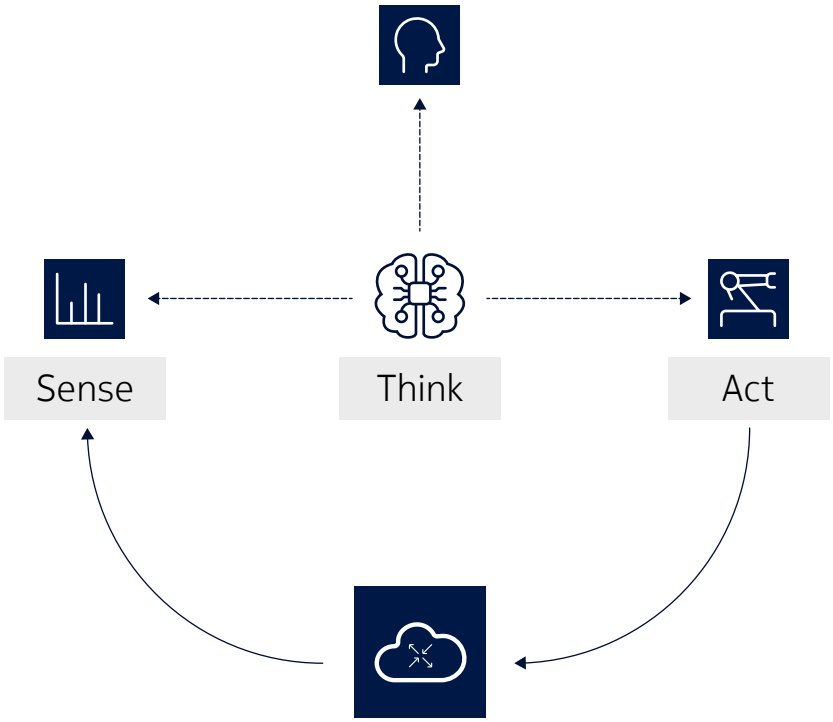
Source: TM Forum

The role of AI in realizing autonomous networking

A fully autonomous network needs to be able to process and make sense of massive volumes of data from many sources, including product documentation and data collected from the network. In the world of network operations, operators are moving towards a closed-loop system where networks can sense, think and act. AI can analyze data from a multitude of sources and help network operators make informed decisions and automate actions within the network. This isn't just about processing data. It's about transforming data into actionable insights, thereby enhancing the agility and responsiveness of communication networks.

When we talk about the benefits of AI and automation, it's essential to recognize the synergy between human ingenuity and technological advancement. Automation helps people and systems execute tasks with unprecedented speed and precision, reducing the margin for manual errors. AI takes this a step further by speeding up processes and making them smarter. The result is a network that's efficient, reliable and adaptable, and capable of delivering personalized services that meet the evolving needs of network operators.

Figure 2: Role of AI in enabling closed-loop systems

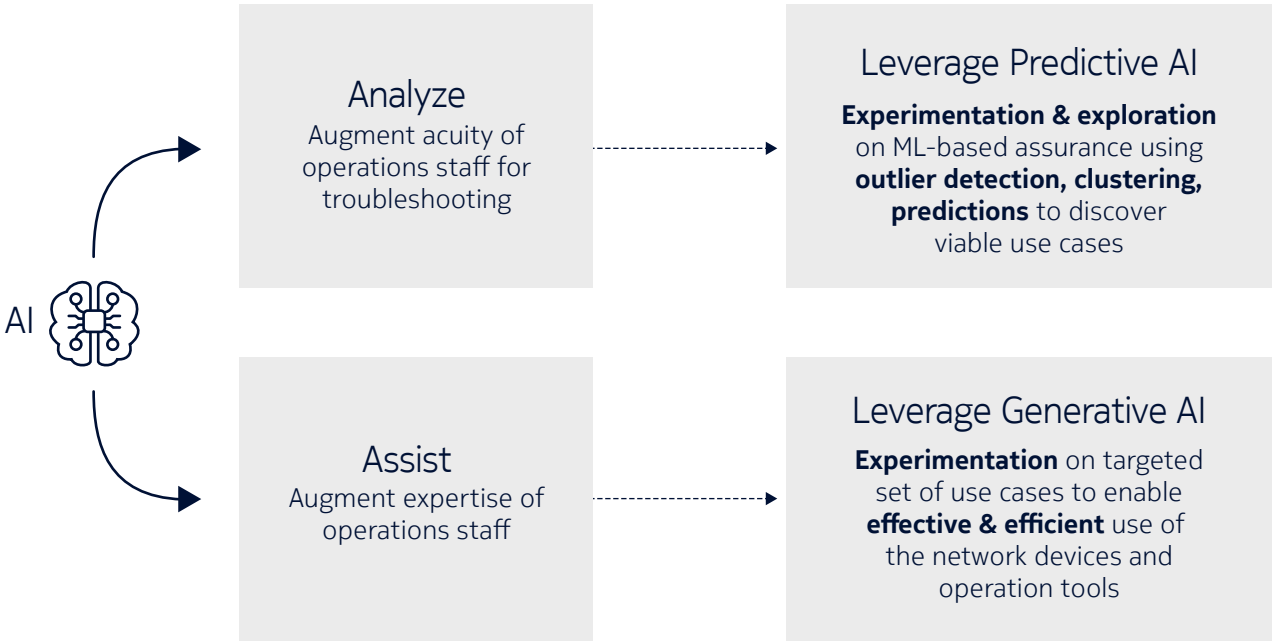


- **Sense:** Collect, store and curate huge variety and volume of data to determine what is happening in the network.
- **Think:** Help humans understand the full complexity of the network with essential cognitive ingredients such as AI and machine learning, which provide unique insights and next best recommendations.
- **Act:** Use closed-loop automation to promptly turn insight into action.

AI operations in IP and data center networks

AI can provide operators with benefits across the network's full lifecycle, from design and deployment to management, optimization and even prediction. It serves a dual purpose for the management of IP and data center networks by supporting predictive AI (classical AI and machine learning) and generative AI.

Figure 3: Dual purpose of AI for network operations



Predictive AI

Predictive AI and machine learning augment the capabilities of the operations team by helping with complex data analysis and decision-making. This is possible with "traditional" machine learning algorithms that have been in use for a few years. Classical AI can quickly analyze vast data streams collected from the network to help identify anomalies and predict and mitigate potential disruptions before they escalate. AI-driven analytics augment human intelligence to improve reasoning for business decisions and accelerate troubleshooting if issues arise. These capabilities improve network performance and increase customer satisfaction.

Generative AI

Generative AI provides assistance to the operations teams to enhance their interactions with the network and with network operation tools. Through advancements in generative AI and large language models (LLMs), operation tools can become more intuitive to lower barriers to entry for users. An AI assistant that uses LLMs to interact with the user in natural language has a comprehensive knowledge of products' capabilities and can provide immediate recommendations on how to use them. It also has a view of the network state and can provide contextual suggestions for troubleshooting or optimizing the network. Finally, it redefines the human–computer interface to simplify complex tasks such as generating software artifacts and customized dashboards.

Predictive AI concepts and use cases

In modern networks, the volume of data (which can include events, alarms and telemetry counters) available to operations teams is increasing at a very fast rate. A network incident can cause hundreds or thousands of events to appear on an operator's dashboards. This is too much information for humans to analyze in a short period of time.

Network operators may also be distracted by events that are unrelated to the problems they are trying to triage. Some problems might not be unearthed until they impact subscribers, and the mean time to identify and resolve a problem will grow and contribute to a general increase in operating cost. Traditionally, operators have used pre-written rules to filter unrelated events and present the meaningful events in a way that helps them identify the root cause of the problem and make decisions faster. However, there are some drawbacks with this method:

- The rules depend on the behavior of the specific network and are not applicable to all environments.
- Manually configured threshold-crossing alerts (TCAs) take effort to write because operators must know the rules in advance and be able to express them.
- The rules can be challenging to manage at scale. The number of rules will increase dramatically as operators grow their networks and deploy new equipment.
- The rules often produce many false positives because they don't reflect dynamic network behavior.

Predictive AI, also known as classical AI, and machine learning offer a way to focus attention on important and relevant events by working out some or all the rules on the operator's behalf, which eliminates the need for the operator to write and maintain them. The operator can rely on these capabilities to automatically organize, correlate and reduce the incoming data, and provide usable information about real problems in the network and how to solve them.

Predictive AI and machine learning is useful for tasks that are too complex for humans to code directly. Some tasks are so complex that it is impractical, if not impossible, for humans to work out all the nuances and code for them explicitly.

A better choice for operators is to let a machine learning algorithm examine a large amount of data and search for a model that will fit. Unsupervised learning is used to correlate data events as the network evolves over time. The information used to train the algorithm is neither classified nor labeled. There is no target or outcome variable to predict or estimate. The algorithm uses the information to cluster events in different groups for specific intervention. It learns more by analyzing data from examples, direct experience or instruction to look for patterns and make better decisions in the future.

Operators can reinforce learning over time by exposing the algorithm to an environment where it can learn from past experience and try to capture the best possible knowledge to make accurate business decisions. The primary aim is to allow the algorithm to learn automatically and adjust instructions accordingly without human intervention or assistance. The machine learning algorithm needs to perform several tasks, as described below.

Baseline training

The first task that the machine learning algorithm performs is to discover the normal performance of the network, called the baseline. To do so, it ingests static and dynamic network data and computes the baseline behavior of the network over a period of several weeks. Operators can accelerate this phase if they have previously collected statistics that can be used to train the system. The detection mechanism

is dynamic, and baselines evolve automatically as the network and traffic evolve. In other words, the expected range around the baseline as computed by the detector rule is adjusted at every interval. Changes in trends are learned automatically and there is no administrative burden to change static thresholds as with simpler thresholding applications.

Anomaly detection

Once the algorithm establishes accurate baselines, it gains the ability to identify meaningful deviations by comparing the actual network behavior against the baseline. Each deviation indicates a potential anomaly. For example:

- Higher-than-normal link utilization can indicate impending congestion.
- An unexpected drop in traffic can indicate silent failures that impact services but don't generate alarms on physical equipment.
- An increase in errors on an interface may indicate hardware issues.

Each newly identified anomaly is an incrementally derived event that can enrich the information provided by existing static and dynamic data sources and enable further correlation and insight. Anomalies are raised immediately when there is a deviation from the baseline. This early warning mechanism can predict failures before they occur.

Event correlation

After detecting an anomaly, the algorithm works in real time to correlate related events (including dynamic data and anomalies) into "incidents" based on patterns. It establishes the relationships between the dynamic temporal-based data and considers rich static data (e.g., service and physical topology) to accelerate the correlation and increase its accuracy.

Operators can then review the incidents and investigate the nature of the problem (e.g., a port misconfiguration, a hardware problem, sunspots that are affecting the microwave network). Once the operator identifies the problem, it can tag and categorize the incident. This is the active learning phase, where the algorithm queries the "teacher" to gain the ability to match a particular incident.

The solution can filter unwanted noise and not present it to the operator. This filtering enables the operator to focus on real problems and root causes and ignore side effects. After resolving an incident, the operator should tag it to describe the action that fixed the problem (e.g., reapply correct configuration, replace card, reset port, redirect traffic).

Corrective recommendation

Once the operator has tagged enough incidents and resolutions, the algorithm can start to make recommendations based on past situations. As the algorithm detects new incidents, it can make recommendations regarding their root cause and the action required to correct them. It makes these recommendations by calculating similarities to past incidents based on a distance score. The operator should evaluate these recommendations and provide feedback about their accuracy. The algorithm uses this feedback to learn how to make more accurate recommendations for future incidents.

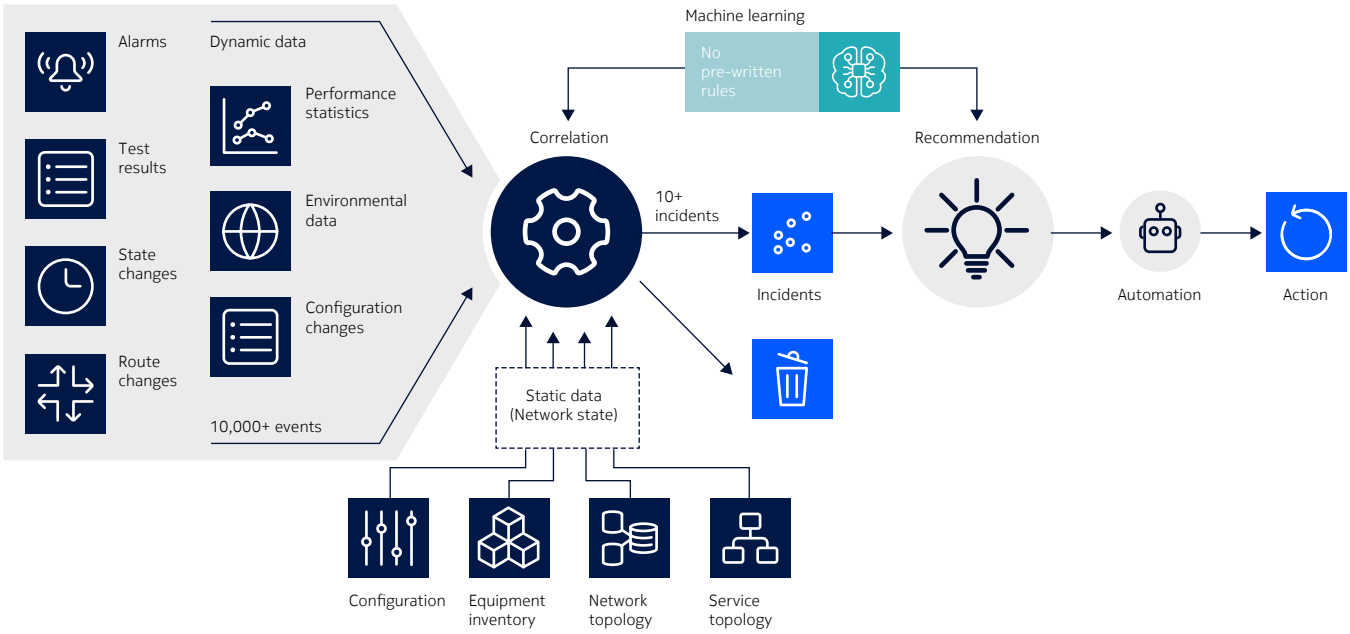
Automated resolution

Once the algorithm has had sufficient training time and the operator trusts the accuracy of its recommendations, it can start to automate the resolution of some types of incidents without human intervention. Automation is particularly useful for incidents that occur regularly over time.

Automated problem resolution brings several benefits:

- It saves time by fixing issues in seconds or minutes instead of hours or days.
- It improves network availability by reducing human errors and service downtime and delivering consistent, predictable outcomes.
- It saves money by cutting the cost of repetitive tasks and making complex operations simple.

Figure 4: Predictive AI use cases for network operations



Predictive AI and machine learning can provide great value to operators by automatically detecting and resolving certain types of issues. The following sections describe some of the use cases that can benefit network operations.

Use case: Fiber cuts

A fiber cut typically generates hundreds of alarms, which makes it difficult for the operator to identify the source of the problem. AI and machine learning can help filter these alarms, clarify the type of incident, pinpoint its location in the network and identify the impacted nodes and services.

Use case: Power outages and device failures

Like fiber cuts, power outages and device failures are problems at physical layers that propagate to upper layers. They cause service disruptions and generate many alarms. AI and machine learning can efficiently cope with these incidents in the same ways they handle fiber cuts.

Use case: Device temperature and power consumption

AI and machine learning can cluster different network resources according to their temperature and look for outliers or for multiple clusters with significantly different temperatures.

Outliers and cluster variances may represent resources that are operating inefficiently, are at risk of premature aging, have improperly installed network devices or have environmental issues.

Use case: Bad packets

AI and machine learning can also cluster ports on network devices according to their speed and the number of bad packets they have received.

Ports normally have zero, or close to zero, bad packets received. However, setting a static threshold may result in a TCA flood for false negatives. Outliers at each port speed might reflect issues with a fiber, transceiver or another device.

Use case: Utilization

Utilization of different network resources such as physical links can also be clustered by AI and machine learning to identify links with significantly higher or lower utilization than their peers. Lower than expected utilization may represent a traffic black hole.

Use case: SLA degradation (latency/throughput)

AI and machine learning can cluster TWAMP Light sessions for various network resources and their latencies by populated slot count. They can also detect outlier resources with higher latency, especially with a trend of increasing or sustained higher latency.

A latency value of zero may indicate a communications or operations, administration and maintenance (OAM) test failure.

Use case: Configuration anomalies

Many network incidents are caused by misconfiguration. AI and machine learning can help by clustering specific types of network attributes, such as the maximum transmission unit (MTU) value set on a port, and by detecting outliers with unexpected values that could be the result of a misconfiguration.

Use case: Silent incidents

Some incidents that impact services are called “silent incidents” because they don’t generate alarms on the physical equipment. This is typical for problems that occur on devices outside the monitored network. An operator may not notice the resulting degradations. AI and machine learning can help identify silent incidents that occur repeatedly. This allows the operator to mark these incidents as known issues.

Use case: Intermittent link errors (link flaps)

In certain weather conditions, third-party leased lines or microwave links can generate intermittent errors called link flaps or toggling links. These errors can cause the IP network to lose router protocol synchronization, which will generate many alarms that then disappear when the link is back up. Recurring flaps can be challenging to investigate and resolve. AI and machine learning can correlate a set of events that indicate a link flap and determine the severity of its impact. They can then automatically notify the link provider and provide some information about the circuit, impact and possible resolution.

Use case: Maintenance windows

Operators plan network updates and upgrades during maintenance windows, typically at night. These activities generate alarms that the operators should ignore or mark with the appropriate tag. This is challenging because a huge number of alarms can be generated during a short time frame. AI and machine learning can group all events related to a maintenance window, making troubleshooting unnecessary. After the maintenance window, the operator can clear all the raised alarms. This eliminates a significant effort that is normally required to ensure flawless network operation after a maintenance event.

Generative AI concepts and use cases

Generative AI models learn patterns from data and can generate new data in response to user queries, called prompts. Generative AI encompasses models that generate images, videos, speech, music and text. Current progress in AI has been largely driven by advances in language modeling. A model approximates a real-world concept or phenomenon. A language model approximates human language and is built through training with a large body of text. This training imbues the model with various properties of language, including aspects of grammar (syntax) and meaning (semantics).

LLMs are machine learning models developed to perform text generation tasks, among others. They enable computers to process, interpret and generate human language to enable more effective human–machine communication. To do this, LLMs analyze or train on massive volumes of text data and thereby learn patterns and relationships between words in sentences. A variety of data sources can be used for this learning process.

LLMs are typically used by network operators to perform the following tasks:

- **Text classification**, which involves categorizing input text into predefined groups. This includes, for example, sentiment analysis and topic categorization. Operators can use sentiment analysis to understand customers' opinions about their services. Email filtering is an example of topic categorization in which email can be put into categories such as "Personal," "Social," "Promotions," and "Spam."
- **Automatic translation** of text from one language to another. Note that this can include areas such as translating code from one programming language to another, such as from Python to C++.
- **Question answering**, which involves understanding and answering questions based on a given text. For example, an online customer service portal could use an NLP model to answer frequently asked questions about a product.
- **Text generation**, which involves generating coherent and relevant output text based on a given input text or prompt.

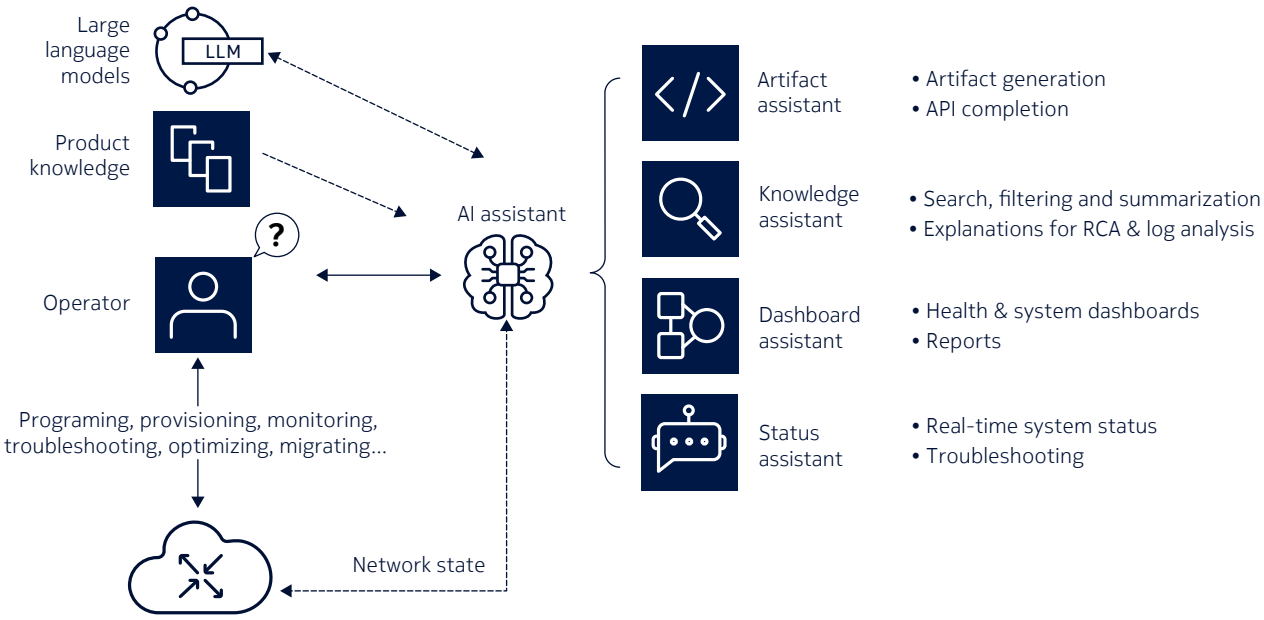
The dynamic nature of data traffic and the complexity of modern networks demand advanced, intelligent solutions that can ensure efficient network operations. Network automation systems, while effective in making networks more responsive and reliable, are often sophisticated and require significant expertise to handle, especially when dealing with heterogeneous network environments. LLMs are proficient at handling natural language interactions. They can simplify tasks for networking teams and provide better experiences for human operators by allowing them to use natural language to interact with management and automation platforms, routers and switches. These capabilities help network operators achieve greater efficiency by using their network operations tools more effectively.

This can enable several key use cases, including providing easy ways to browse documentation, check device and network state, troubleshoot issues and query configuration examples.

As illustrated in Figure 5, an AI assistant intercepts the message based on keywords contained in the question and collects contextual information to augment the data to be processed by the LLM. This information could include live network state data from the management system or the network or parts of the product documentation.

The assistant then sends an API call to the LLM with a question along with any associated contextual information that will enable the LLM to analyze the question more deeply. The LLM receives this request and processes the question and associated information. It then returns a natural language response back to the assistant, which in turn shows the response to the user.

Figure 5: Generative AI use cases for network operations



The following sections describe some of the use cases that can leverage LLMs for contextualized communication and guided actions to assist the network operators.

Use case: Artifact assistance

Coding assistance can be embedded into the developer environment to help network operators with software artifact generation, explanation and elimination of human errors due to typos, syntax mistakes and logical defects. These capabilities can decrease inefficiencies and lengthy debugging processes and increase productivity and democratize the use of automation. Artifacts could include scripts, workflows, device adaptors, intents and APIs. For example, a network developer can ask AI to generate a new intent for deployment of a new service type between two sites.

Use case: Knowledge assistance

Generative AI can support intelligent product assistance by processing product documentation and best practices to quickly provide the most relevant information to the operators. As an example, a network operator can ask AI about a certain routing protocol or how to upgrade a system.

Use case: Dashboard assistance

Generative AI can provide aid and explanation that helps operators create health dashboards and reports. One example could be that the operator needs a custom visualization of the overall bandwidth utilization in the network and asks AI to help build a dashboard that can provide it.

Use case: Status assistance

Generative AI can also enable operations teams to use natural language to interact with network operation tools and devices to monitor their status in real time, troubleshoot and take actions if necessary. For example, the network operator can ask AI to check the system health or identify the device with the highest number of alarms in the network.

Challenges

The promise of LLMs is immense. However, there are challenges with this technology, including privacy, hallucination, overreliance and vendor lock-in. There are ways to mitigate these challenges.

Data privacy and security

Leveraging AI's capabilities entails responsibilities, especially regarding data privacy. To function effectively, AI models require access to vast amounts of data, which is one of the most valuable assets for operators. Exposing this sensitive data to AI models can cause some concerns.

The sanctity of network data is crucial. Several avenues can be explored to mitigate these concerns:

- Stringent data governance agreements with cloud providers of the AI systems ensure that when data does move to the cloud, it is treated with the utmost care and respect for privacy.
- Data anonymization ensures that data carries no sensitive information as it feeds into AI systems.
- For those preferring closer control, on-premises solutions and using open-source models locally can help with privacy and data security where user data is sensitive and offer a way to realize the benefits of AI while keeping data firmly within their domain.

Hallucination

Fine-tuning and retrieval-augmented generation (RAG) are effective strategies for reducing hallucination in generative AI models, where the model produces incorrect or fabricated information.

- Fine-tuning includes training a preexisting AI model on a specific dataset relevant to networking. This process improves the AI model's understanding of the network, which enables it to create more accurate outputs that are less prone to hallucination. By tailoring the model to a specific context, fine-tuning reduces the likelihood of generating incorrect or out-of-context information.
- RAG combines a generative model with an external knowledge retrieval system. It retrieves relevant, up-to-date information from a database or external sources and uses this data to inform the generation process. This approach helps to keep the model's responses in line with facts and verifiable data to minimize hallucinations by anchoring its outputs to reliable sources.

Together, the two strategies significantly reduce hallucinations. Fine-tuning improves the model's expertise in networking, while RAG ensures that responses are based on accurate, real-time information.

AI vendor lock-in

In a rapidly evolving industry, reliance on one specific vendor's model or platform limits flexibility and adaptability. New AI models with superior performance, efficiency or capabilities are continuously being developed. If a system is locked into a single vendor, it may miss out on other innovations, leading to suboptimal outcomes or increased costs.

It is crucial to build solutions that are flexible and open to integrating different AI models. Through compatibility with diverse models and platforms, network operators can easily adopt new and better-performing models as they emerge. This adaptability ensures long-term success, cost-efficiency and the ability to leverage the best AI technologies.

Conclusion

While fully autonomous networks remain a future goal, integrating AI into networking operations marks a significant leap in network management and optimization. [1]

The integration of natural language processing (NLP) related to artificial intelligence for network operations (AIOps) is evolving. Networking-related AIOps capabilities may be enabled through an operations support system (OSS), business support system (BSS) or network management and automation platform, or directly integrated into a networking hardware platform such as an IP router or Ethernet switch. The end goal is to apply AI-based approaches to improve the human operator experience and enhance the day-to-day processes best suited to the network operator's unique business needs.

Nokia is very excited to be part of this journey. We are integrating AI more extensively into our network automation platforms for the IP domain through the Nokia Network Services Platform (NSP) and the data center domain through the Nokia Event-Driven Automation (EDA) platform. The GenAI assistant application built into the Nokia SR Linux network operating system for routers and switches can also benefit the networking teams and help improve the human operator experience by enabling them to use natural language to interact directly with the router or switch when needed.

We are also working with our customers to evaluate and exploit the benefits that AI capabilities will bring to their existing systems as well as the way these capabilities will change the way humans interact with these systems.

As we venture further into this AI-driven era, we maintain an unwavering commitment to data privacy and security. This aspect of AI is of utmost importance to its success and it is our highest priority.

References

[1] Converge! Network Digest and AvidThink, “2024 AI in Networking – Pipe Dreams and AI Realities,” [Online]. Available: https://nextgeninfra.io/2024-ai-networking/?spotlight=nokia-cory. [Accessed 13 October 2024].

Abbreviations

AI	artificial intelligence
AIOps	artificial intelligence for IT operations
API	application programming interface
EDA	Event-Driven Automation
GenAI	generative AI
IP	Internet Protocol
LLM	large language model
MTU	maximum transmission unit
NLP	natural language processing
NOS	network operating system
NSP	Network Services Platform
RAG	retrieval-augmented generation
TCA	threshold crossing alert
TMF	TM Forum
TWAMP	Two-Way Active Measurement Protocol

About Nokia

At Nokia, we create technology that helps the world act together.

As a B2B technology innovation leader, we are pioneering networks that sense, think and act by leveraging our work across mobile, fixed and cloud networks. In addition, we create value with intellectual property and long-term research, led by the award-winning Nokia Bell Labs.

With truly open architectures that seamlessly integrate into any ecosystem, our high-performance networks create new opportunities for monetization and scale. Service providers, enterprises and partners worldwide trust Nokia to deliver secure, reliable and sustainable networks today – and work with us to create the digital services and applications of the future.

Nokia OYJ
Karakaari 7
02610 Espoo
Finland
Tel. +358 (0) 10 44 88 000

Document code: CID214310 (October)