GREYSKIES

THE FUTURE OF TELECOM SERVICE ASSURANCE: ENABLING THE JOURNEY TO AUTONOMOUS NETWORKS

GreySkies Inc.

4111 E Madison Street
Seattle, WA 98112, USA
https://www.greyskiesinc.com/

Copyright © 2025 GreySkies, Inc. All rights reserved.

Table of Contents

Executive Summary

The telecommunications industry is undergoing a significant shift as 5G, IoT, and cloud-native technologies introduce new levels of network complexity. This shift is happening while operators must continue to manage the myriads of legacy networks and technologies. Traditional service assurance models, which are often reactive, siloed, limited in scope, adaptability and reconfigurability can no longer meet the demands of modern networks. Operators now require a proactive and intelligent framework that provides real-time insights, unified visibility, and automated remediation across hybrid and multi-vendor environments.

This whitepaper introduces a next-generation service assurance strategy that transforms how operators monitor, analyze, and act on network behavior. It is built on five core pillars: comprehensive data collection, observability, AI-driven analytics, closed-loop automation and context-aware AI Assistant. Generative AI further enhances operations by enabling conversational interfaces, intelligent workflows, and predictive insights that simplify tasks and accelerate resolution. We also touch on the role of emerging advances and use cases of Agentic AI within Service Assurance.

With this model, telecom operators can reduce mean time to resolution, lower operational costs, and improve service quality. More importantly, it establishes the foundation for autonomous operations where networks can adapt and optimize with minimal human intervention, enabling operators to move up the Autonomous Level scale as defined by the TM Forum.

Introduction

Service assurance in telecom has traditionally been guided by the FCAPS model; Fault, Configuration, Accounting, Performance, and Security, a framework designed to structure the management of network operations. While this model remains useful, it falls short in modern environments. Traditional tools are often fragmented, reactive, and limited in scope. This siloed approach makes it harder to detect issues early, correlate data across systems, and respond quickly to problems. The result is slower incident resolution, increased operational costs, and a greater risk of service disruptions.

Fully decommissioning existing tools is rarely practical; it is costly, risky, and often unnecessary. Furthermore, modernizing the FCAPS framework is not something that can happen overnight, it is a gradual process, as each underlying platform undergoes modernization or sunsetting, which takes time. A more pragmatic approach is to layer an intelligent, cross-domain platform on top of the current framework. This allows operators to protect their existing investments while providing immediate benefits from visibility, correlation, and coordination of the modern platform. Over time, as the retained FCAPS components are modernized, they are already embedded within a more advanced and integrated service assurance environment.

The evolution toward more advanced and complex network infrastructures necessitates a shift away from traditional, FCAPS-centric OSS frameworks. Modern service assurance platforms must leverage technologies such as artificial intelligence (AI), automation, and robust data pipelines to achieve proactive network management and enhanced operational efficiency. These capabilities empower operators to move from reactive to predictive strategies, ensuring seamless service continuity and rapid response to potential issues.

This white paper introduces a next-generation service assurance model based on observability, intelligent automation, AI-driven analytics, and Generative AI. It explains how telecom operators can improve visibility, reduce operational overhead, respond to issues more quickly, and prepare for the transition to more autonomous network operations.

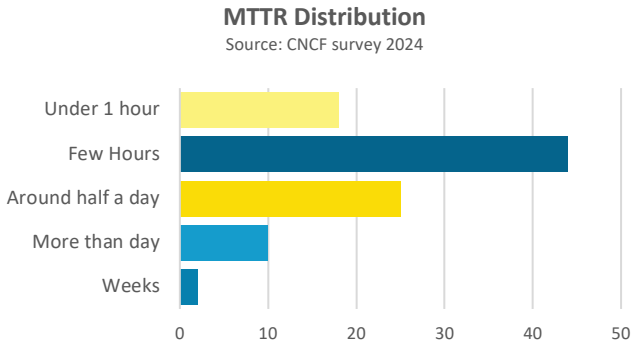
Limitations of Traditional Service Assurance

Siloed Systems and Fragmented Data

Traditional service assurance tools are often deployed in silos, with each system monitoring a specific domain or function. This separation limits the ability to share data across systems and reduces visibility into the full network environment. As a result, operators struggle to build a unified view of service performance, which slows down troubleshooting, increases operational costs, and complicates coordination between teams.

Across the industry, operators frequently encounter situations where a single underlying issue triggers multiple alarms across domains, yet no system links them all together. This lack of cross-domain correlation leads to delays in identifying root causes, especially during time-sensitive incidents.

These limitations are reflected in real-world performance metrics. As shown in the chart below, a significant portion of operators still report mean time to resolution (MTTR) ranging from several hours to more than a day. This delay is largely driven by fragmented monitoring systems and the absence of intelligent, unified workflows.



Reactive Management

Most legacy service assurance systems are fundamentally event-driven and threshold-driven, designed to raise alarms only when a metric crosses a static limit, or a device reports a fault. This architecture means they cannot anticipate or prevent issues, they only respond once service degradation is already underway. Because they rely on static rules and signatures, they lack the ability to detect emerging or unknown patterns, and operators are forced into constant manual tuning. The result is a system that recognizes only the problems it has been explicitly programmed to catch, and always after the fact.

Compounding this, these platforms remain resource-centric rather than service-centric. They generate floods of raw, device-level alarms without translating them into customer or service impact, leaving operations teams to triage symptoms instead of addressing root causes. NOCs become trapped in a fire-fighting cycle, waiting for alarms, opening tickets, and reacting to incidents, rather than proactively safeguarding service quality. This operational model entrenches a reactive posture, where issues are managed after they occur, rather than predicted, prevented, or automatically mitigated.

Operational Impact of Fragmented Systems

When systems are not integrated, operations teams spend more time switching between tools, correlating data manually, and responding to redundant or irrelevant alarms. Because each domain (RAN, transport, IP, core, IT, etc.), relies on its own tools and data models, NOC engineers must swivel across multiple consoles and manually piece together events to diagnose issues. This siloed visibility slows root cause analysis, inflates mean time to repair, and leaves operators chasing symptoms rather than understanding the end-to-end service impact. A common metric that reflects this inefficiency is ticket re-assignments, defined as the number of times a ticket is handed off between teams before resolution. In fragmented environments this number is often high, signaling costly inefficiency, longer outages, and poorer service quality.

Fragmentation also amplifies noise and misalignment. Each domain system generates its own raw alarms, often duplicating or overlapping symptoms of the same underlying fault. Without correlation, NOCs are inundated with redundant or low-value alerts, fueling fatigue and raising the risk that genuine incidents are missed or misclassified. At the same time, service-level visibility remains elusive: operators cannot easily determine whether customers are affected, making it harder to prioritize and act decisively. This lack of coherence entrenches inefficiency and perpetuates a cycle of firefighting, where tickets are repeatedly bounced across siloed teams, and service quality is managed reactively rather than proactively.

Data Collection and Integration Gaps

Traditional assurance platforms were built for static, domain-specific networks, consequently they struggle to collect and normalize data across today's complex, multi-layer, multi-vendor environments. This approach tends to require extensive human intervention and creates partial insights and critical blind spots: infrastructure-level data may be visible, but service-level or customer-experience signals are often missing, or vice versa. Without seamless, dynamic integrations across network, service, and experience layers, operators cannot reliably determine the true impact of issues, nor can they prioritize actions based on customer or business criticality. The result is a reactive mode of operation where teams chase alarms rather than managing outcomes.

These gaps grow even more severe as operators transition to virtualized and cloud-native architectures. Legacy tools were never designed to instrument dynamic elements like VNFs, CNFs, or containerized workloads, which spin up and down fluidly across distributed infrastructures. Entire parts of the service chain may remain invisible, creating monitoring "black holes" that

increase operational risk and prolong incident resolution. In such environments, the inability to stitch together data across physical, virtual, and cloud-native domains not only delays root-cause analysis but also undermines confidence in service assurance itself, making it harder for operators to deliver consistent, high-quality customer experiences.

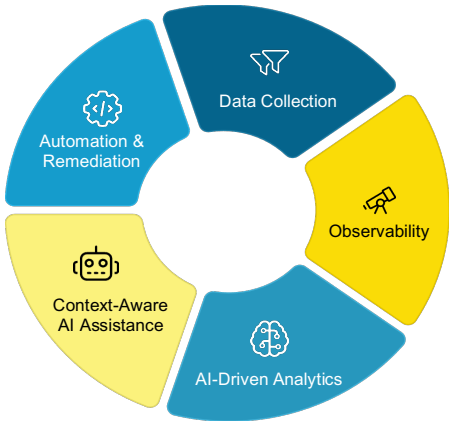
Next-Generation Service Assurance: A Holistic Approach

As operators adapt their networks to support cloud-native architectures, automation, and real-time operations, expectations for assurance systems are also evolving and elevating. The shift from static, device-centric infrastructures to dynamic, software-defined, and distributed environments has dramatically increased complexity and volatility. Traditional device-focused models built to monitor fixed hardware and simple thresholds are no longer adequate to capture the fluid behaviors of virtualized functions, containerized workloads, and multi-cloud service chains. At the same time, the business priorities of operators are shifting whereby assurance is no longer seen as a back-office monitoring function but as a strategic enabler for customer experience, operational efficiency, and faster service innovation. This redefinition of assurance is prompting both operators and technology vendors to re-evaluate its role and requirements in the modern telco stack.

According to Gartner's Market Guide for AI Offerings in CSP Network Operations (2025), the direction of the service and network assurance market is being shaped by increasing interest in AI/ML, Generative AI, AIOps, and automation, particularly as they apply to service and network assurance.

The shortcomings of legacy assurance have made one thing clear: operators can no longer rely on siloed, device-centric tools that react only after problems manifest. Modern networks (cloud-native, software-defined, service-driven, etc.) demand a new foundation for assurance, one that is proactive, intelligent, and end-to-end. A next-generation assurance platform must not only detect faults more quickly, but also understand their impact, predict their likelihood, and guide operators toward faster, automated resolution. In short, assurance must evolve from a back-office monitoring function into a strategic enabler of resilient operations and superior customer experience.

At the heart of this evolution is a holistic framework built on five interdependent pillars. Each pillar addresses a critical gap left by legacy systems: from ensuring complete and trustworthy data collection, to introducing observability techniques that reveal dynamic service behaviors; from applying AI-driven analytics that detect anomalies and anticipate issues, to embedding automation and closed-loop remediation that shorten resolution times; and finally, to providing context-aware AI assistance that empowers operations teams with actionable insights in real time. Together, these pillars create a foundation for assurance that is predictive rather than reactive, service-aware rather than device-focused, and agile enough to support the rapid pace of network and service innovation.



5 Pillars of Next Generation Service Assurance

Comprehensive Data Collection

Next-generation service assurance begins with the ability to gather data from across the entire operator ecosystem. This includes infrastructure, networks, services, applications, and user experience metrics. To support advanced analytics, automation, and proactive monitoring, operators must adopt a data collection model that is both broad in scope and high in quality.

Collect Everything

Modern networks are highly distributed, spanning physical infrastructure, virtualized network functions, and containerized workloads running in public and private clouds. Service assurance platforms must be able to ingest telemetry, logs, metrics, and traces from all these sources. This includes access, core, and transport networks, as well as service platforms such as IMS, policy engines, and control-plane functions.

The "collect everything" model emphasizes the importance of not leaving any data blind spots. It covers

diverse sources such as event logs from virtual network functions, KPIs from radio access nodes, customer experience probes, configuration changes, SLA violations, and even contextual signals from the business support layer. These data points form the backbone of observability, helping operators monitor service delivery, identify early degradation, and predict network behavior before issues arise. From telemetry to policies, from probes to platforms, if it has a digital interface, it should be part of the assurance fabric.

Data Quality

The value and impact of any AI model, dashboard, or automation workflow depends entirely on the quality of the data it is based on. Poor-quality data can trigger false alarms, hide critical incidents, and lead to failed remediations. In contrast, clean, enriched, well-structured data enables accurate insights, reliable trend forecasting, and confident decision-making.

To ensure this level of reliability, next generation service assurance platforms must embed quality controls throughout the data pipeline. During ingestion, validation checks confirm that records are complete, well-formed, and accurately timestamped. Once accepted, the data is normalized into consistent formats across vendors and enriched with critical metadata such as service types, customer segments, device classifications, and geographic locations.

Key attributes of high-quality data include:

- **Accuracy**, reflecting true operational conditions.
- **Completeness**, ensuring full coverage across domains.
- **Timeliness**, to support real-time decision-making.
- **Consistency**, to unify metrics from different systems.
- **Contextual richness**, to enable business-aligned insights.

To improve operational trust, it is recommended that platforms assign quality scores or confidence tags to data streams, allowing downstream processes to factor in data reliability when making decisions.

Data Collection and Integration

A comprehensive data strategy must be supported by seamless integration capabilities. This involves collecting telemetry, KPIs, logs etc. from the network, exchanging data with orchestration and configuration systems, and exposing insights to upper OSS and business systems.

Effective architectures support three main integration flows:

Southbound: Ingests telemetry from physical, virtual, and cloud-native elements.

East-West: Connects with orchestration, inventory, and policy systems to contextualize and enrich data.

Northbound: Delivers processed insights to reporting tools, SLA dashboards, and customer experience systems.

By ensuring that data can flow freely and reliably between layers, operators eliminate information silos and create a single source of truth for service assurance.

Data Enrichment and Standardization

Once data is collected, it must be enriched with additional context to make it meaningful and actionable. This includes enriching raw telemetry with inventory metadata, applying standardized dictionaries and ontologies, and running quality checks to ensure accuracy and consistency. Enrichment is what transforms disparate, low-level signals into structured insights that reflect the true state of services, resources, and customer experience. Without it, operators are left with fragmented data streams that are difficult to interpret or correlate across domains.

Standardization is equally critical. By harmonizing data into a common model, operators eliminate silos and create a single, trustworthy backbone for assurance. This unified foundation not only supports consistent analytics and reporting but also accelerates the application of AI/ML, which depends on clean, normalized inputs. Together, enrichment and standardization close a key gap left by legacy systems and lay the groundwork for AI-driven operations that scale with modern network demands. This ultimately lays the foundation for assurance systems to perform cross-correlation and “stitch” together true end-to-end views, enabling proactive network health management, faster root-cause analysis, and prioritization based on actual service impact.

Observability

Comprehensive Visibility for Enhanced Network & Service Performance

Observability is a cornerstone of next-generation service assurance, extending far beyond the limits of traditional monitoring. Monitoring tells you when something breaks; observability explains why it is happening and how it affects the service chain. Traditional monitoring is limited to tracking predefined metrics or thresholds at the device level, while observability provides real-time, end-to-end visibility across infrastructure, services, and customer experience. In today's highly dynamic and distributed networks, this shift allows operators to move from fragmented snapshots of individual devices to a holistic understanding of service health, system behavior, and user impact.

Legacy monitoring tools were designed for static environments and typically focus on isolated elements or domains. They generate siloed dashboards and disconnected alerts that fail to capture how issues cascade across the service chain. Modern observability platforms, by contrast, ingest and correlate data from multiple layers such as spanning infrastructure, applications, and user sessions, to create a unified, contextual view. This cross-layer visibility enables operators to trace the ripple effects of a single fault, such as a transport bottleneck or RAN issue, all the way through to its impact on latency, video quality, or call completion rates.

The value of observability becomes especially clear during large-scale, high-traffic events, such as national holidays, major sporting matches, or festivals, when network behavior is unpredictable and highly dynamic. By surfacing patterns and detecting anomalies as they emerge, observability equips operators to intervene proactively, making real-time adjustments that preserve performance and ensure customer experience remains seamless.

Turning Data into Actionable Insights

Observability is not just about collecting metrics. It is about transforming raw data into meaningful, actionable insights. This requires advanced data modeling, contextual enrichment, and real-time analysis. Operators must be able to move from "What happened?" to "Why did it happen?" and "What should we do next?", all within a single workflow.

Modern platforms achieve this through dynamic dashboards, contextual correlation, and built-in intelligence. When an anomaly is detected, observability systems help trace its impact across services, identify affected users, and provide historical context to determine whether this is an isolated case or part of a recurring pattern.

This level of insight dramatically reduces the time needed to detect, triage, and resolve issues. It also supports better communication across teams by presenting unified, easily understandable views of network health.

AI-Driven Analytics

As networks become more distributed, dynamic, and software-driven, the volume and velocity of assurance data have grown beyond the capacity of human operators and traditional rule-based systems.

AI-driven analytics equips operators not only to detect problems as they emerge, but also to connect seemingly unrelated signals across domains, predict outcomes, and recommend the best course of action. This approach moves assurance from being a passive monitoring function to an intelligent, decision-support engine that scales with network complexity. Within this paradigm, three areas stand out as foundational: real-time anomaly detection to surface issues before they impact users, AIOps correlation to connect the dots across noisy data streams and reveal true root causes, and decision support systems to guide operators with contextual, AI-powered recommendations. Together, these capabilities redefine assurance as a predictive and service-aware discipline, reducing outages, accelerating resolution, and enabling operators to run networks with greater confidence and efficiency.

Real-Time Anomaly Detection

AI has become integral to modern service assurance, with machine learning (ML) models enabling real-time anomaly detection across network, service, customer, and business domains. Unlike static thresholds or manual monitoring, ML algorithms can identify subtle deviations and emerging patterns as they occur, allowing operators to catch issues before they escalate into service-impacting incidents. This not only accelerates response times but also improves the precision of root-cause analysis, ensuring service quality is protected even in complex, multi-domain environments.

One of the most immediate benefits is within an operator's NOC/SOC. Today, engineers spend long hours manually scanning dashboards, tracking KPIs, and attempting to spot abnormal behavior in critical metrics. Once anomalies are noticed, they are passed downstream for root-cause analysis and remediation, a process that is slow, error-prone, and inconsistent due to its reliance on human judgment. By contrast, ML-based anomaly detection provides a fast, consistent, and scalable way to surface abnormal patterns in near real time. This reduces outages, shortens mean time to detect (MTTD), and frees engineers to focus on resolution and optimization rather than repetitive monitoring tasks.

It is important to note that the enormous volumes of data generated across modern digital services present a significant challenge for ML-based anomaly detection. While many machine learning algorithms can detect irregular patterns in controlled or small-scale settings, these approaches often become computationally prohibitive when applied to telecom operator environments that must monitor thousands of concurrent data streams. Simply having a "good" algorithm is not enough, such algorithms must be engineered to be computationally efficient, so they can process high-throughput data in real time, and scalable, so they can adapt as the system grows without degradation in performance. Equally important, they must be robust and reliable, capable of handling incomplete, noisy, or imperfect data while continuing to provide meaningful insights. In practice, this means anomaly detection systems must be designed holistically, coupling algorithmic efficiency with distributed and fault-tolerant infrastructures, ensuring that they not only detect anomalies accurately but also sustain that performance at telecom operator scale under dynamic and imperfect real-world conditions.

AIOps Correlation: Connecting the Dots Across Domains

AIOps correlation enables telecom operators to move beyond reactive incident handling by intelligently linking related events, metrics, and logs across network domains. In traditional assurance systems, each element generates isolated alarms, leaving operators to manually piece together the puzzle of what is actually happening. This often leads to alert fatigue, redundant tickets, and delays in identifying the true root cause. AIOps addresses this challenge by applying machine learning techniques such as clustering, graph analysis, and causal inference to unify noisy, multi-domain data streams into coherent incident narratives.

For example, a sudden latency spike at the service layer can be automatically traced to congestion in the transport network or excessive CPU utilization on a virtualized function. Instead of overwhelming the NOC/SOC with hundreds of raw alarms from routers, switches, or VNFs, AIOps correlation consolidates them into a single, prioritized incident enriched with context such as topology relationships, service dependencies, and historical patterns. This unified view accelerates root cause analysis, supports automated remediation, and facilitates cross-team coordination by ensuring all stakeholders are working from the same source of truth. One operator described this capability as a shift "from alarm storms to incident stories."

As networks continue to scale in complexity spanning physical, virtual, and cloud-native domains, the role of AIOps correlation becomes indispensable. It not only lowers mean time to detect (MTTD) and mean time to resolve (MTTR) but also reduces operational overhead by filtering noise and minimizing unnecessary escalations. Ultimately, AIOps correlation transforms incident management from a manual, error-prone process into a data-driven, intelligent workflow that improves service performance, reduces OPEX, and enhances customer satisfaction.

Decision Support Systems

Beyond day-to-day incident management, AI-driven analytics also play a pivotal role in long-term strategic decision-making. Telecom operators face choices around where to invest, how to allocate scarce resources, and how to balance performance, cost, and sustainability. Decision Support Systems (DSS) powered by AI bring structure and intelligence to these choices by synthesizing massive volumes of cross-domain data into clear, actionable recommendations.

For example, AI-powered DSS can deliver value across multiple domains, including:

- **Network Optimization**: ML models continuously analyze utilization, congestion, and fault patterns to identify where capacity upgrades are needed most, preventing both over-investment and bottlenecks.
- **5G Rollouts:** AI prioritizes deployment areas based on demand forecasts, spectrum efficiency, and revenue potential, ensuring infrastructure investments deliver maximum return.
- **Energy Efficiency:** AI models trade off power consumption against quality of service to uncover opportunities to reduce energy use without

degrading customer experience, aligning operations with sustainability goals.

By moving beyond static reports and manual forecasting, AI-powered decision support provides operators with a forward-looking, data-driven compass. It not only enhances planning accuracy and resource utilization but also strengthens the link between operational insights and executive strategy, ensuring telecom networks evolve in line with both business and environmental objectives.

Balancing Cost and Operational Effectiveness

AI-driven analytics provides flexibility for both immediate operational needs and long-term resource planning.

- **Real-Time Analytics**: Enables rapid responses to anomalies or network changes, minimizing downtime and ensuring continuous service quality.
- **Scheduled and Historical Analytics**: Analyzes trends and patterns over time, uncovering opportunities for cost optimization and process improvements.

This dual approach helps operators balance operational efficiency with cost management. By integrating real-time anomaly detection, cross-domain insights, and strategic decision-making, telecom operators can optimize networks dynamically, enhance service quality, and support long-term strategies, including sustainability and technological advancement.

Automation and Remediation

As networks scale in size and complexity, manual intervention in assurance workflows becomes increasingly unsustainable. Traditional incident management depends heavily on human operators to triage alarms, perform diagnostics, and trigger corrective actions, a process that is slow, resource-intensive, and prone to error. In modern, cloud-native and distributed environments, where issues can propagate across layers in seconds, this reactive model is no longer sufficient to safeguard service quality.

Next-generation assurance platforms embed automation and remediation at the core of their design, shifting from human-driven processes to intelligent, closed-loop responses. By codifying best practices, leveraging AI/ML insights, and integrating with orchestration and control systems, automation ensures that many issues are resolved before they impact end users. At the same time, it augments human operators with actionable playbooks and guided interventions, reducing both downtime and operational overhead.

To structure this evolution clearly, we can think of automation and remediation in four complementary dimensions. First are **self-healing mechanisms**, the "immune system" of the network that automatically recovers from common faults. Second is the **reduction of manual intervention**, where automation absorbs repetitive operational tasks, eliminating inefficiencies such as ticket reassignments and inconsistent workflow execution. Third comes **actionable automation**, where detection seamlessly translates into remediation through predefined or adaptive playbooks. Finally, the **customer impact** dimension demonstrates why these capabilities matter, by directly improving uptime, service continuity, and trust. Viewed together, these four perspectives describe not just a set of features, but an integrated framework that shifts assurance from reactive firefighting to proactive, resilient operations.

Self-Healing Mechanisms

Automation is transforming network management by embedding self-healing workflows that minimize downtime and ensure service reliability. Instead of waiting for human intervention, these workflows can automatically initiate corrective actions such as opening support tickets, running diagnostic tests, applying targeted configuration updates, and performing post-remediation validation checks. The goal is to restore service quality seamlessly and consistently, while reducing the burden on operations teams.

A critical safeguard in these workflows is the use of rollback mechanisms, which allow systems to revert to the last known stable state if validation steps fail. This ensures that attempted fixes never worsen the situation, and that resilience is maintained even under unpredictable conditions.

More advanced implementations extend beyond reactive healing to predictive self-healing, where machine learning models anticipate anomalies before they manifest into service disruptions. By learning from historical patterns and real-time telemetry, these systems can trigger preventative actions, such as traffic rerouting, capacity adjustments, or pre-emptive configuration changes, effectively shifting operations from reactive firefighting to proactive assurance.

Minimizing Manual Intervention and Enhancing Efficiency

Large-scale telecom operations still rely heavily on manual tasks that consume engineer time and introduce inconsistency. Automation addresses this by reducing reliance on human effort for repetitive and error-prone activities such as root-cause triage, ticket creation and routing, compliance checks, and configuration audits. By standardizing these processes into automated workflows, operators achieve both greater efficiency and more consistent execution.

For example, when a core network element fails, automation can immediately identify the impacted services, consolidate alarms, and route the ticket to the correct resolver group, thereby avoiding the costly cycle of reassignments that delays resolution. Similarly, automated compliance audits ensure that configuration baselines are continuously validated without requiring teams to run manual checks. These efficiencies not only reduce mean time to detect and repair, but also free skilled engineers to focus on higher-value activities such as performance optimization, service innovation, and long-term capacity planning.

Impact on Downtime and Customer Satisfaction

The ultimate measure of automation's value is its impact on customers. By accelerating resolution times and preventing incidents from escalating, automation reduces mean time to detect (MTTD) and mean time to resolve (MTTR), two of the most important drivers of customer experience. Self-healing mechanisms prevent many issues from ever reaching the end user, while proactive monitoring ensures that potential degradations are addressed before they disrupt service.

For customers, this translates to fewer dropped calls, smoother video streams, and more reliable connectivity. These outcomes directly influence satisfaction, trust, and loyalty. For operators, it means improved SLA compliance, reduced churn, and stronger brand differentiation in highly competitive markets. In this way, automation is not only an operational efficiency enabler but also a customer experience catalyst, turning resilience into a competitive advantage.

Reducing Downtime Through Actionable Automation

Automation in assurance goes beyond generating insights, it drives closed-loop action. Once anomalies are detected, predefined or dynamically generated playbooks can be triggered to remediate the issue immediately. This can include rebalancing workloads across VNFs and CNFs, scaling cloud-native resources, adjusting routing policies, or reallocating spectrum during periods of high demand. By integrating directly with orchestration and control layers, actionable automation ensures that detection is seamlessly connected to remediation.

The result is a shift from "monitor and notify" to "detect and act". Rather than relying on humans to interpret dashboards and initiate fixes, automation operationalizes best practices in real time, reducing downtime and preserving service quality. Over time, as systems learn from repeated scenarios, these playbooks evolve from predefined scripts to adaptive, intent-driven actions, accelerating the path toward fully autonomous operations.

Context-Aware AI Assistance

As telecom operations become more data-intensive and fast-moving, teams are increasingly looking for AI tools that can support decision-making, accelerate analysis, and reduce the time needed to act. Context-aware AI assistants offer a new layer of intelligence, helping engineers and analysts navigate complex data, summarize incidents, and even co-create solutions through conversational interfaces.

To fully leverage AI assistants, the next generation service assurance platform, must be built on a flexible architecture that can evolve with new technologies. At the core is a unified data fabric that ingests and correlates diverse operational data such as alarms, logs, metrics, traces, and topology, so AI assistants have consistent, high-quality context. On top of this, prompt management and context-intelligence layers translate user intent into grounded, reliable outputs while maintaining awareness of services, incidents, and operator preferences. The platform must support dynamic model selection (locally hosted / cloud-based), routing tasks between general-purpose LLMs, domain-specific models, and analytics engines as needed, all while scaling in real time to handle high-volume data during outages. Governance, security, and trust are critical, ensuring explainability, access controls, and compliance are built in. Finally, the architecture must

enable human-in-the-loop validation, allowing engineers to refine AI outputs and continuously improve relevance. With these characteristics in place, service assurance platforms can seamlessly integrate AI into operations, unlocking the full potential of incident summarization, conversational data insights, and AI co-pilots for automation design.

Incident Summarization with Operational Context

When incidents occur, it is critical to understand what happened, where, why, and what needs to be done. Traditionally, this involves sifting through logs, alarms, and metrics, often across multiple systems. Context-aware AI assistants can dramatically shorten this process by generating real-time, human-readable summaries of major events.

These assistants correlate alarms, extract relevant trends, and map events to affected services or locations. The result is a timeline or narrative that helps teams grasp the situation at a glance. Rather than reading hundreds of log lines, an operator can see a concise summary outlining the root cause, progression, and suggested next steps.

This improves situational awareness, enables faster handovers between shifts, and supports clearer post-incident reporting and audit trails.

Chat with Data for Instant Insights

Beyond summarization, AI assistants can act as conversational interfaces to the network's operational data. Engineers can ask natural-language questions such as "Which sites had the highest packet loss yesterday?" or "Show me the services impacted during the last major outage" and receive immediate, contextual answers.

These assistants are powered by large language models trained to understand network-specific terminology and workflows. They can navigate logs, dashboards, and telemetry streams to extract relevant information, often accompanied by visualizations or links to deeper insights. This reduces time spent searching for data and allows teams to shift focus from discovery to action.

AI Co-Pilot for Design, Build, and Validation

Context-aware AI is also transforming how telecom teams build and maintain analytics and automation use cases. Engineers can now collaborate with AI co-pilots to design, implement, and test new workflows through an interactive development experience.

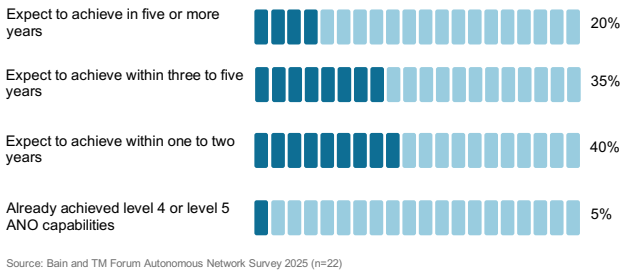
For example, an operator can describe a desired use case such as detecting subscriber-level QoE degradation or automating a capacity check, and the assistant can suggest data sources, draft analytic logic, generate code in Python or SQL, and validate output against sample data.

This accelerates prototyping, lowers the barrier to automation adoption, and enables teams to focus on refining outcomes rather than starting from scratch. The assistant can also help with documentation, error checking, and integration with broader service assurance workflows.

Enabling Autonomous Networks

The capabilities of next-generation service assurance, highlighted in the previous section, form the operational backbone for realizing Autonomous Networks as envisioned by the TM Forum. The TM Forum defines Autonomous Networks (AN) as self-configuring, self-optimizing, and self-healing systems that minimize human intervention while delivering consistent, high-quality services. To achieve this vision, operators must advance through the well-known AN maturity model (Levels 0-5).

Next-generation assurance platforms provide the building blocks required for operators to climb this ladder. Data collection and enrichment eliminate blind spots and create a unified, trusted data fabric. Observability delivers real-time visibility into service behavior across layers. AI-driven analytics detect anomalies, correlate events, and produce insights at machine speed. Automation and remediation close the loop by turning those insights into guided or autonomous actions that reduce downtime and protect services. And with context-aware AI assistance, human operators are augmented with real-time recommendations, accelerating the journey to higher levels of autonomy.



Source: Bain and TM Forum Autonomous Network Survey 2025 (n=22)

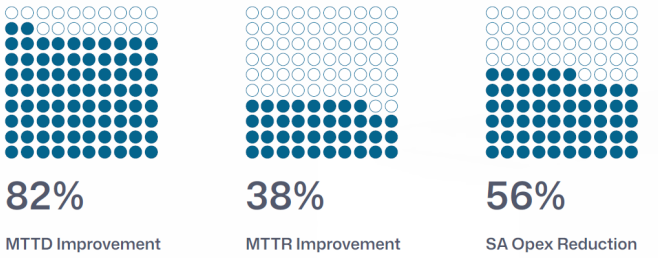
Industry progress, however, shows that most operators remain early in their AN journey. According to Capgemini Research Institute and TM Forum (2024), 84% of CSPs are still at Level 1 or Level 2, with 61% targeting Level 3 within the next five years. Bain & Company and TM Forum's 2025 survey of global operators suggests incremental progress, with around 20% of CSPs now achieving Level 4 maturity in select domains, and an additional 35% expecting to reach Level 4 or 5 within two years. These findings highlight both the urgency and opportunity: while ambition is high, the majority of operators are still working to embed the foundational capabilities of service assurance needed to progress beyond assisted and partial autonomy.

By layering these capabilities, operators can move progressively along the TM Forum maturity model from assisted operations (L1) to conditional autonomy (L3) and ultimately high autonomy (L4–L5). In practice, this means networks that dynamically adapt to changing conditions, protect customer experience without waiting for manual intervention, and optimize resources with minimal OPEX overhead. In short, next-generation service assurance is not just an incremental upgrade, it is the enabler of Autonomous Networks, turning the concept from aspiration into operational reality.

Operational Benefits of Assurance Modernization

With the GreySkies Service Assurance Platform deployed across leading telco operators, service providers have achieved improvements in operational efficiency, service reliability, and responsiveness. By leveraging AI-driven analytics, real-time anomaly detection, and intelligent automation, they improved mean time to detect (MTTD) by as much as 82%, enabling faster issue identification and empowering operations teams to take proactive measures to safeguard subscriber experience. They also realized up to a 38% reduction in mean time to repair (MTTR), minimizing service disruptions and strengthening SLA compliance.

Many of these operators also experienced operational expenditures reduction for Service Assurance by up to 56%, driven by tool consolidation, automated workflows, and more efficient fault and performance management processes.

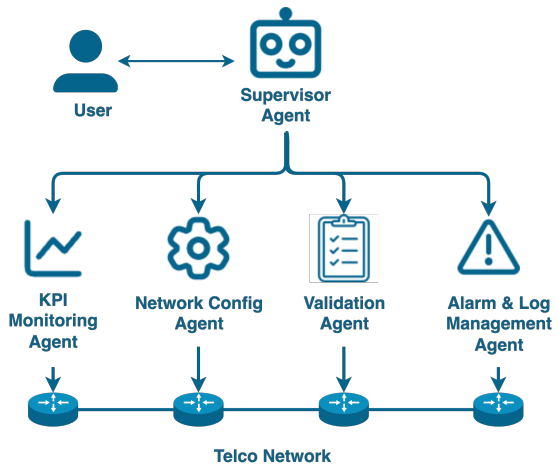


What's Next - Agentic AI Service Assurance

Agentic AI represents the next major evolution beyond Generative AI and AI Assistants. While Generative AI focuses on content creation and AI Assistants support task execution through human prompts, Agentic AI goes further by introducing autonomous, goal-driven agents that can independently plan, decide, and act to achieve defined outcomes. These intelligent agents are not limited to static instructions. They combine capabilities like long-term memory, contextual understanding, adaptive planning, and continuous feedback loops to operate with minimal human intervention.

For telecom operators, this shift is transformative. Agentic AI can intelligently manage and optimize complex, large-scale network operations, customer interactions, and service provisioning in real time. By continuously learning and adapting, these agents can respond to changing network conditions, preempt issues before they escalate, and drive operational efficiency. This results in faster service delivery, reduced operational costs, and an enhanced customer experience.

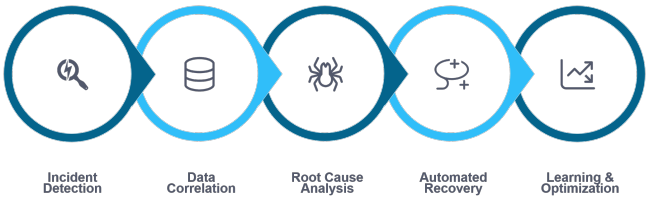
While operators are in the early phases of adopting AI within their operations, they are steadily progressing toward identifying the needs, opportunities, and practical use cases for Agentic AI as the technology matures.



Below are potential use cases where Agentic AI can deliver significant value for telecom operators.

Closed-Loop Incident Management

In incident remediation workflows, Agentic AI can autonomously manage the full lifecycle from detection to resolution. Upon identification of an anomaly or performance deviation, agents ingest telemetry data across domains such as fault, performance, and topology, correlate it against historical patterns, and identify the probable root cause. Rather than waiting for manual intervention, the agent can initiate recovery steps or trigger automated scripts through integrated orchestration systems. Over time, it can learn from prior outcomes and optimize its decision models, improving the accuracy and efficiency of responses.



Autonomous NOC Agents

Within network operations centers, Agentic AI can serve as always-on agents monitoring streaming telemetry, logs, and real-time events. Unlike traditional alert-based monitoring, these agents proactively identify anomalies, perform contextual correlation across domains, and determine whether intervention is required. Based on prior patterns and operational policies, they can either notify operators with recommended actions or initiate pre-approved mitigation flows. Over time, they adapt thresholds and refine detection logic, reducing alert noise and false positives.

Energy Efficiency and Sustainability Agents

Agentic AI can support sustainability initiatives by optimizing the energy usage of service infrastructure. These agents monitor performance, utilization, and environmental variables, dynamically shifting workloads, disabling underutilized resources, or rerouting traffic to more energy-efficient paths. In multi-cloud or distributed networks, they may recommend off-peak scheduling or regional consolidation of workloads to reduce overall energy consumption. These actions are made within service constraints to preserve SLAs and quality. The result is a reduction in carbon footprint without compromising operational goals, supporting green telecom initiatives.

Autonomous RAN Configuration and Optimization

Managing radio access networks (RAN) involves balancing a large number of parameters, such as power control, modulation schemes, scheduling, and handover thresholds, which directly affect service quality and spectrum efficiency. These parameters are highly dynamic, requiring frequent adjustments in response to traffic patterns, user mobility, and environmental conditions. Traditional rule-based optimization methods, while effective for static scenarios, cannot adapt quickly enough to the complexity and pace of modern 5G and beyond-5G networks.

Agentic AI addresses this challenge by introducing continuous, autonomous configuration and optimization across the RAN. Through real-time monitoring of performance metrics like signal-to-noise ratio, bitrate, and retransmission rates, agentic agents can identify suboptimal configurations and apply targeted adjustments. These agents' reason through trade-offs such as maximizing throughput while minimizing interference, and coordinate parameter changes across network layers and cell sites. By automating this process, agentic AI reduces operational overhead, improves response times, and enhances spectrum utilization. It enables RAN operations to shift from manual, reactive tuning to autonomous optimization based on real-time network conditions.

Regulatory and Compliance Monitoring

Regulatory compliance in telecom networks often requires constant monitoring of service behavior against defined standards or internal policies. AI agents can be configured to scan logs, SLA reports, and configuration states for potential violations. For example, they can detect data retention anomalies, exposure of protected subscriber information, or SLA breaches tied to contractual commitments. When conditions deviate from acceptable limits, the agent triggers alerts or initiates mitigation actions as permitted. These agents operate continuously and evolve to recognize new compliance risks, reducing the manual effort and lag time associated with audit-based methods.

Conclusion

As telecom networks evolve to meet rising demand, increasing complexity, and customer expectations, traditional approaches to service assurance are no longer sufficient. Fragmented tools, reactive workflows, and siloed data make it difficult for operators to maintain service quality, resolve incidents quickly, or scale operations effectively.

This paper has outlined a modern, AI-driven approach to service assurance one built on five key pillars: comprehensive data collection, deep observability, intelligent analytics, automated remediation, and context-aware AI assistance. Together, these capabilities enable operators to shift from reactive troubleshooting to proactive, data-informed decision-making.

By investing in platforms that unify insights across domains, apply real-time intelligence to vast datasets, and automate both diagnostics and response, telecom operators can improve operational agility, reduce mean time to resolution, and strengthen customer experience. Context-aware AI assistants further amplify this impact by empowering teams to engage with data more intuitively and build advanced use cases more rapidly. These capabilities provide the foundation on which operators' journey towards TM Forum Autonomous Networks Levels 4-5 are built.

The future of service assurance lies in combining human expertise with intelligent systems that are scalable, adaptable, and fully aligned with the needs of next-generation networks. Operators that make this shift will be better positioned to manage complexity, optimize performance, and deliver reliable, high-quality services in a fast-changing digital landscape.