RADCOM

White Paper

The Single Source of Truth

Unified Service Assurance Data Layer Powers AI Agents on the Path to Autonomous, Customer-Centric Networks

Introduction

The telecommunications industry is approaching a critical inflection point. The rapid acceleration of artificial intelligence (AI) coincides with an unprecedented surge in network demand, driven by the proliferation of connected devices and a 32% increase in global mobile traffic in 2025, according to IDC. These forces are reshaping the scale, complexity, and expectations placed on telecom networks.

Operators are now tasked with managing these highly dynamic, multi-layered environments while striving to deliver greater efficiencies, increased automation, and consistently high service quality, all under sustained pressure to keep costs low. Simultaneously, they must contend with vast volumes of network telemetry and behavioral data, creating both an opportunity and a challenge: the ability to interpret, prioritize, and act on data in real time.

In response, operators are turning to advanced AI initiatives, including agentic AI, to drive a new era of operational intelligence. This is shifting AI from a “bolt-on” capability to a foundational layer in telecom operations, reshaping how networks are designed, managed, and optimized, and enabling the transition toward truly autonomous, data-driven networks. However, achieving true automation is operationally complex. It requires a fundamental shift from fragmented data silos to a unified, trusted data foundation that delivers real-time, end-to-end visibility and actionable insights. Such a foundation enables the aggregation, correlation, and analysis of data across services, transforming raw information into a strategic asset that drives operational efficiency and competitive differentiation.

With this foundation in place, the operational model evolves further: from managing isolated AI tools to orchestrating autonomous, agent-driven systems. In this paradigm, AI agents operate across vendor-agnostic environments, leveraging shared data to perceive network conditions, reason about root causes, and take proactive corrective action.

This white paper examines how to enable this transformation through next-generation AI-native assurance solutions. It describes a new operational model where AI agents, supported by a unified data layer, continuously monitor, analyze, and respond to network conditions in real time, addressing issues before they affect customer experience and ultimately enabling fully autonomous networks.

> Achieving true automation requires a fundamental shift from fragmented data silos to a unified, trusted data foundation

Background: State of the Autonomous Network

The concept of autonomous network operations has been formalized by the TM Forum through its Autonomous Networks (AN) initiative, which defines a maturity scale from Level 0 (fully manual) to Level 5 (fully autonomous, no human intervention required). According to the recent NVIDIA State of AI in Telecommunications: 2026 Trends Survey Report, 88% of organizations are between levels 1-3 of autonomy, with 17% at Level 1, 35% at Level 2 and 36% at Level 3. These levels include conditional automation within defined domains, and with human oversight for cross-domain decisions. The industry target, however, is Level 4 by the end of this decade with the aim of achieving intent-driven, largely self-managing networks that escalate to humans only for novel, high-impact situations.

How do operators plan to get to Level 4 network automation?

41% of respondents expect to leverage agentic AI to drive this transition, according to the NVIDIA survey. In parallel, 3GPP has established complementary architectural foundations through the Network Data Analytics Function (NWDAF), first introduced in Release 15 and significantly expanded in Releases 16 and 17. NWDAF provides a native 5G framework for collecting, analyzing, and distributing network data to enable automated, data-driven decision-making.

The convergence of these approaches highlights a common principle: autonomous operations depend on high-quality, comprehensive, real-time data. Without such a foundation, AI agents are forced to operate on incomplete or delayed information, resulting in unreliable outputs and reduced trust in automation.

This is particularly critical in the context of AI inference, i.e., the process by which trained models generate predictions and insights from live data. The accuracy and timeliness of inference are directly tied to the quality and completeness of the underlying data.

TM Forum Autonomous Network Maturity Levels

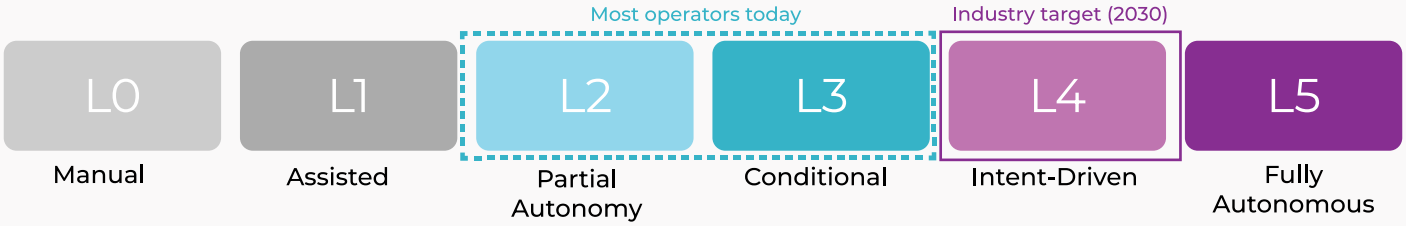


Figure 1: TM Forum Autonomous Network Maturity Levels, most operators sit at L2-L3 today

Source: TM Forum Autonomous Networks Initiative

However, industry data shows that this remains a significant challenge with 27% of respondents identifying data-related issues as the leading barrier to autonomous networks in NVIDIA's survey. A separate GSMA Intelligence survey, conducted in partnership with RADCOM, reflects a similar concern; only 39% of respondents are fully prepared to handle future data workloads for autonomous networks.

This underscores a fundamental reality: while AI models and architectures continue to advance, the path to Level 4 and Level 5 autonomy ultimately depends on delivering a robust, real-time data foundation that supports reliable, continuous decision-making at scale.

Barriers to Autonomous Network Operations

The gap preventing operators from moving from today's Level 2-3 operations to the autonomous network vision is not primarily a model problem. It is a data problem. Below are three specific barriers that prevent operators from deploying AI agents with the confidence needed for autonomous action.

> Only 39% of respondents are fully prepared to handle future data workloads

Source: NVIDIA State of AI in Telecommunications 2026 | GSMA Intelligence / RADCOM Survey

Figure 2: Three data gaps blocking the path to autonomous network operations

1. Incomplete Visibility into Subscriber Experience

Conventional service assurance and network analytics architectures are fundamentally network-centric, focused on monitoring infrastructure-layer KPIs such as radio access network (RAN) availability, interface throughput, and core network function health. While these metrics provide insight into component-level performance, they lack true end-to-end service visibility across the full user plane and control plane interaction path. More importantly, they do not deliver granular, per-subscriber observability, limiting the ability to accurately assess real-time Quality of Experience (QoE) at the individual session level.

In practice, this creates **a disconnect between network health indicators and actual service experience.** For example, during a degraded video streaming session, infrastructure key performance indicators (KPIs) may remain within nominal thresholds, masking impairments occurring at the application or transport layers. Without correlated, end-to-end telemetry and user-level session data, these degradations remain undetected even as subscribers feel the impact.

Additionally, this challenge is complicated by the way many traditional assurance systems ingest and analyze network data. Instead of analyzing all network activity, they rely on sampling, i.e., capturing only a portion of user sessions. While this approach worked in earlier network generations, it creates significant visibility gaps in today's high-volume, fast-changing 5G environments.

With so many users, devices, and services interacting at once, these gaps mean operators can miss subtle or short-lived issues that often have the biggest impact on the customer experience. As a result, AI systems that rely on this incomplete data are working with only part of the picture. This limits their ability to accurately detect issues, understand root causes, and make reliable decisions.

Without 100% session-level visibility, AI agents inherit the gaps and the operator inherits the complaints. By capturing high-quality data across all user sessions, they can gain full visibility, make better decisions, and ensure a consistently high-quality experience.

2. Siloed Data that Cannot Support Cross-Domain Reasoning

AI agents operating within a single domain are inherently limited to the scope of the data they can access. In practice, **most service degradations are multi-domain in nature, spanning the RAN, transport, core, and service layers**. For example, a subscriber experiencing poor voice quality may be impacted simultaneously by RAN congestion, a misconfigured Quality of Service (QoS) policy in the core, and a routing anomaly within the IP multimedia subsystem

(IMS). An agent constrained to a single domain can identify only a contributing factor and not the true root cause.

This limitation is compounded by fragmented data-collection architectures, in which telemetry is generated, processed, and stored in domain-specific silos (e.g., RAN counters, core KPIs, IMS logs). Each of these has different formats, granularity, and time alignment.

The absence of a unified data layer makes cross-domain correlation complex, resource intensive, and often incomplete. Without consistent, end-to-end data collection anchored in per-subscriber, per-session context, it is not possible to accurately reconstruct service behavior across domains.

Effective root cause analysis, therefore, requires a unified data fabric that spans the full network and service architecture, using subscriber session data as the common correlation key. With this in place, AI agents achieve true cross-domain reasoning and deliver accurate, actionable outcomes.

3. Delays that Defeat Proactive Action

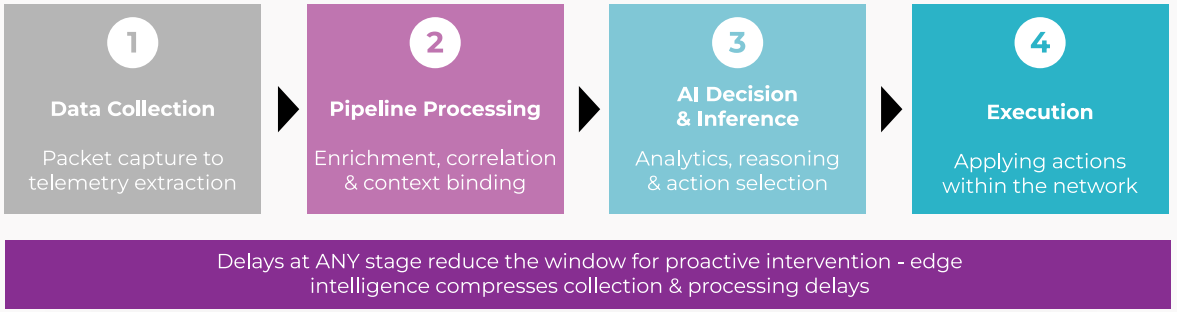


Figure 3: The data-to-action latency pipeline, delays at any stage reduce the window for proactive intervention

A key distinction between reactive and proactive operations lies in the timing and availability of data used for decision-making in autonomous networks. In AI-driven environments, agents must detect early indicators of service degradation at the per-session level and initiate corrective actions before user experience is impacted. This requires continuous, near real-time access to high-fidelity telemetry that can be consumed and reasoned over by AI agents operating within closed-loop assurance frameworks.

Achieving true real-time responsiveness, however, depends on minimizing several distinct sources of latency across the data-to-action lifecycle: data collection latency (from packet capture to telemetry extraction), pipeline latency (including enrichment and correlation), decision latency (analytics and AI inference), and execution latency (applying actions within the network).

While much of the current focus is on reducing pipeline latency, delays in any of these stages can significantly impact the effectiveness of autonomous operations.

Edge intelligence plays a key role in enabling this shift by bringing data processing and insight generation closer to where network events occur. This helps to reduce latency across multiple stages and empowers AI agents with immediate, context-rich visibility. Nevertheless, realizing this potential depends on how efficiently data is captured, processed, analyzed, and acted upon across the entire lifecycle.

However, even with edge intelligence, many existing data pipelines introduce latency that limits the effectiveness of autonomous operations. Batch processing, polling-based data collection, and aggregation at the network function level all delay the exposure of relevant data to analytics and AI systems. As a result, AI agents often operate on delayed or partially aggregated inputs rather than real-time, session-level insights. By the time the data is available for analysis and action, the opportunity for pre-emptive intervention may have already passed. This constrains autonomous networks' ability to act proactively and consistently deliver optimal service outcomes.

Data Foundations for Autonomous Network Operations

A service assurance platform capable of supporting autonomous AI agents must satisfy several foundational requirements. Collectively, these criteria distinguish between AI agents operating on a complete, real-time representation of the network and those operating on partial or approximated views derived from incomplete data.

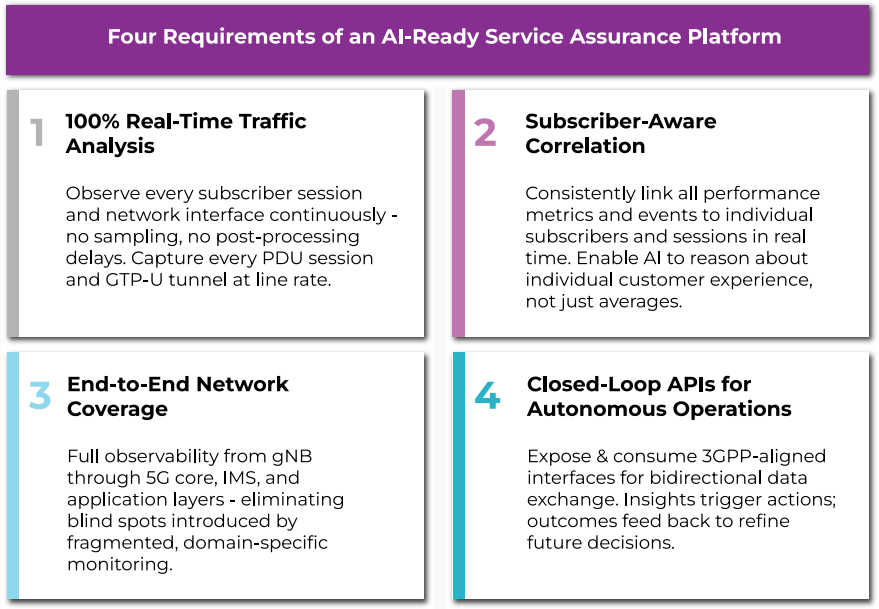


Figure 4: The four requirements every AI-ready service assurance platform must satisfy

1. 100% Real-Time Traffic Analysis:

The assurance platform should observe every subscriber session and network interface continuously, delivering 100% real-time visibility without reliance on sampling, post-processing delays, or coarse aggregation. Traditional approaches that rely on sampled telemetry or periodic polling introduce statistical gaps and latency, limiting visibility into transient behaviors and edge-case degradations. Full-fidelity observability at the session level ensures that every Protocol Data Unit (PDU) session, GPRS Tunneling Protocol – user plane (GTP-U) tunnel, and user-plane transaction is captured and analyzed in real time. With the emergence of distributed compute resources such as Data Processing Units (DPUs) at the network edge, high-throughput packet processing and telemetry extraction can now be performed at line rate, enabling scalable, per-session visibility across 5G networks.

2. Subscriber-Aware Correlation:

All performance metrics and events must be consistently correlated to individual subscribers and their associated sessions in real time. This requires maintaining a persistent context that links network-level telemetry to specific user equipment (UEs), PDU sessions, Quality of Service (QoS) flows, and application transactions. Without this correlation, analytics remain limited to aggregated domain-level KPIs, preventing AI agents from reasoning about individual customer experience. Subscriber-aware data enables AI systems to move beyond averages and distributions to understand the precise impact of network conditions on specific users and services.

3. End-to-End Network Coverage:

The assurance platform must provide observability across the full-service delivery chain, from the next-generation node-B (gNB) in the RAN through the 5G core network functions, and onward to the IMS and application-layer services. This represents true end-to-end visibility aligned with the 3GPP service-based architecture, rather than the post hoc reconciliation of siloed domain metrics. By maintaining continuity of context across control-plane and user-plane interactions, the system enables AI agents to correlate events across domains within a unified data model, eliminating blind spots introduced by fragmented monitoring.

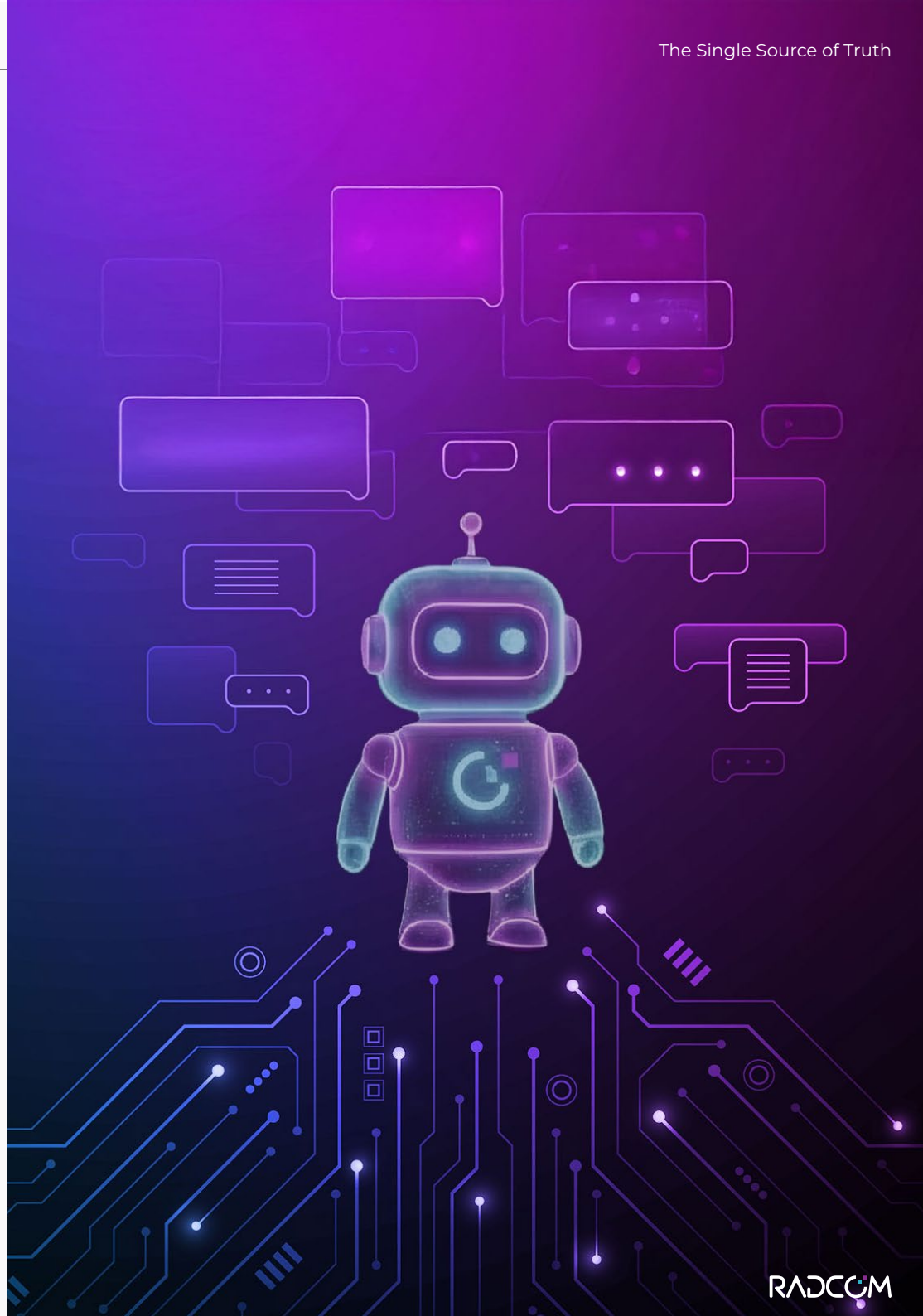
4. Closed-Loop APIs for Autonomous Operations:

To support autonomous operations, the assurance platform must expose and consume standardized, 3GPP-aligned interfaces. This enables seamless integration with AI agents and external analytics systems without prescribing a specific architecture or framework. These interfaces should support both the delivery of analytics (such as insights,

predictions, and detections) and the ingestion of feedback or decisions generated by AI agents. This bidirectional data exchange is essential for enabling closed-loop automation so that insights can trigger actions, while the outcomes of those actions are fed back into the system to continuously refine and improve future decisions.

When these four conditions are met, namely: real-time full-traffic visibility, subscriber-level correlation, end-to-end observability, and NWDAF-aligned closed-loop interaction, then AI agents are equipped with a complete, high-fidelity representation of network and service behavior. This foundation enables them to operate with genuine confidence, supporting accurate reasoning, reliable automation, and effective autonomous operations at scale.

> These four criteria distinguish AI agents operating on a complete, real-time network view from those operating on partial or approximated data



Understanding Agentic AI and the Role of AI Agents

For the purposes of this white paper, Agentic AI refers to artificial intelligence systems designed to operate autonomously, pursuing defined goals and making decisions with minimal human intervention. Unlike traditional AI, which primarily provides insights or recommendations, agentic AI determines and executes the optimal course of action to achieve a desired outcome. Agentic AI represents an architectural paradigm, combining multiple AI models, data sources, and decision-making capabilities into a coordinated system. At its core are AI agents or autonomous entities that perceive data, apply contextual reasoning, and take action. These agents operate in a continuous loop of perception, reasoning, and execution, enabling real-time responsiveness and closed-loop automation.

In the telecom context, agentic AI has the potential to transform networks into truly autonomous systems capable of managing complex operations, understanding dynamic priorities, and proactively optimizing performance. AI agents can ingest data from multiple sources, maintain contextual awareness, and act independently within defined policies and governance frameworks.
These agents do not operate in isolation. They collaborate and communicate with one another to orchestrate multi-step workflows across domains, enabling coordinated decision-making at scale.
This means telecom networks can evolve beyond automation into intelligent, self-operating environments, powered by AI agents that:

- Interpret data and take context-aware action
- Operate autonomously with minimal supervision
- Collaborate across agents to execute multi-step, cross-domain workflows

Operational Domains for AI Agents in Autonomous Networks

With the above data foundations in place, AI agents can be deployed across distinct operational domains, each targeting a different dimension of network performance. Each domain aligns to a distinct set of operator objectives, while all agents operate on a shared underlying data layer that provides consistent, real-time, end-to-end visibility. Within this architecture, agents are not constrained to isolated datasets or domain-specific silos; instead, they leverage the same unified source of truth while applying specialized models, logic, and workflows tailored to their respective functions. This enables coordinated, cross-domain intelligence while preserving domain-specific depth of analysis.

The four operational domains are:

- Customer Experience: Agents focused on monitoring, analyzing, and optimizing Quality of Experience (QoE) at the per-subscriber and per-service level
- Service Quality: Agents that evaluate service-level performance metrics, detect degradations, and ensure compliance with defined service expectations and service level agreements (SLAs)
- Network Optimization: Agents responsible for real-time monitoring, fault detection, root cause analysis, and operational stability across network domains
- AIOps: Agents that automate operational workflows, correlate events across systems, and support closed-loop decision-making and orchestration

Together, these domains form a comprehensive AI agent taxonomy, enabling operators to address customer experience, service assurance, and network efficiency through a unified, data-driven operational model.

Customer Experience

The customer experience domain focuses on individual subscriber outcomes and the real-time assessment of Quality of Experience (QoE) at the per-user level. AI agents in this domain analyze session-level telemetry and correlated network data to determine the current experience of a specific subscriber and identify early indicators of potential dissatisfaction or service degradation. This enables a shift from reactive, ticket-based handling to proactive identification and mitigation of issues before they escalate into complaints.

In addition, these agents support evidence-based validation of reported issues by correlating subscriber complaints with underlying network and service-level metrics. By reconstructing the conditions of a given session, they provide customer care teams with an objective view of what occurred across the network and service layers. Beyond individual cases, this domain also extends to assessing broader subscriber impact. AI agents can determine which users are currently affected by a given issue and estimate the scale of impact if the problem persists. This allows operators to prioritize remediation efforts based on both severity and the number of impacted subscribers, improving responsiveness and overall customer satisfaction.

Service Quality

The service quality domain focuses on ensuring that network and service performance meets defined expectations from an end-to-end perspective. AI agents in this domain validate incidents not only from a network

perspective, but based on actual subscriber experience, confirming that service has returned to normal conditions as perceived at the user level. In this context, resolution is defined by the restoration of acceptable Quality of Experience (QoE) across affected subscribers, rather than by isolated configuration or infrastructure changes.

In addition, agents in this domain perform multi-layer analysis to identify the underlying causes of service degradation. By correlating events and telemetry across the RAN, transport, and core network domains, they generate ranked, evidence-based hypotheses regarding potential root causes. This multi-hop reasoning enables engineering and operations teams to focus on the most probable sources of the issue and take targeted corrective actions with greater speed and precision.

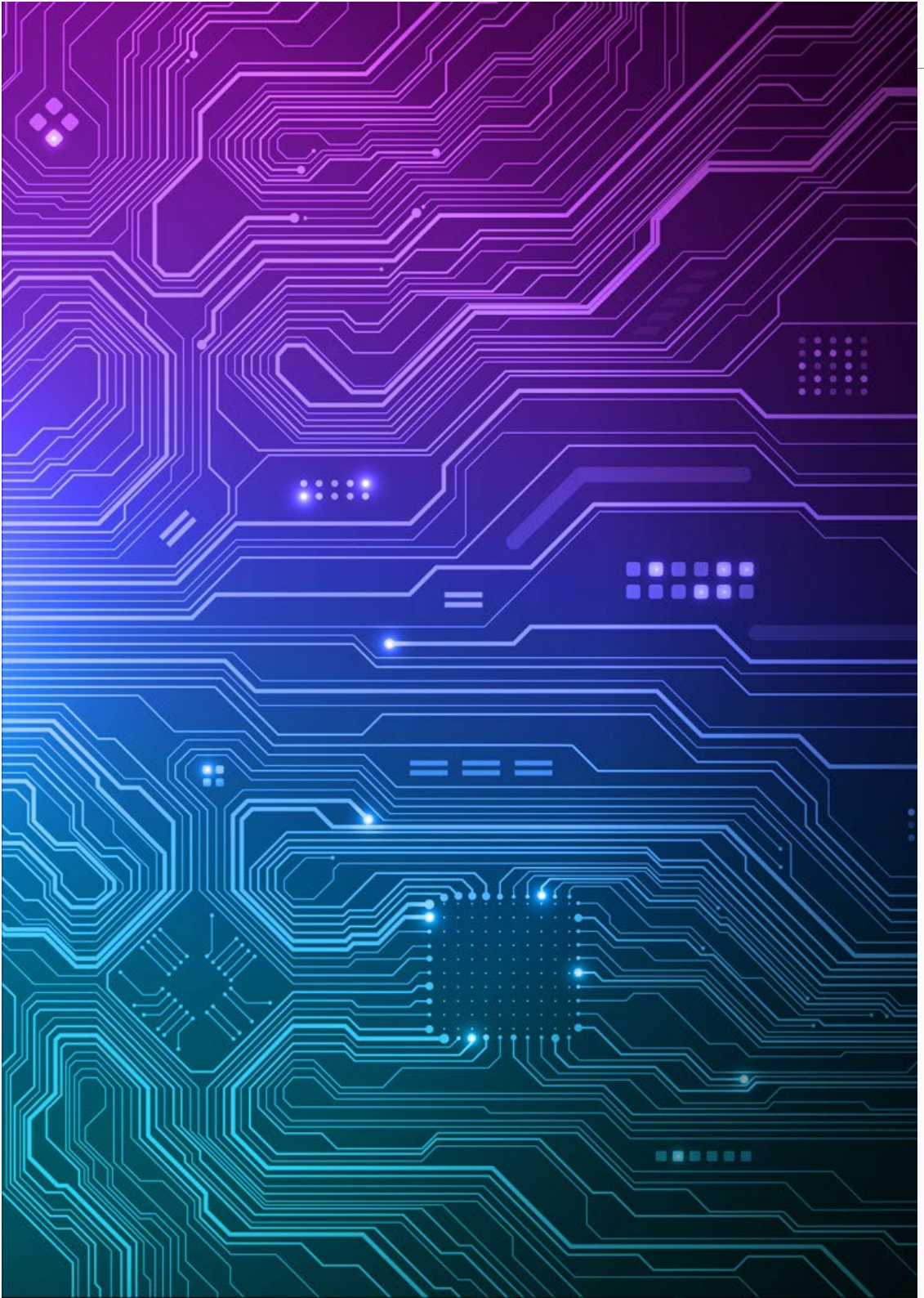
Network Optimization

The network optimization domain focuses on improving network performance, efficiency, and service quality through data-driven recommendations and continuous learning. AI agents in this domain analyze current network conditions alongside historical patterns of previously resolved incidents to generate actionable remediation recommendations. By leveraging a growing corpus of case outcomes, these systems refine their recommendations over time, enabling increasingly accurate and context-aware optimization strategies.

In parallel, agents perform detailed analysis of radio access network (RAN) conditions to identify issues such as coverage gaps, weak signal strength, interference, and mobility or handover inefficiencies. Based on these insights, they provide targeted recommendations to address detected problems, supporting proactive network tuning and optimization. This enables operators to maintain consistent service quality while adapting to evolving traffic patterns and network conditions at scale.

AIOps

The AIOps domain serves as the coordination layer across all operational domains, enabling automation, correlation, and orchestration of insights generated by AI agents. It aggregates findings from multiple sources, correlates events and alerts across domains, and prioritizes actions based on business and operational impact. In addition, it supports the automation of operational workflows by integrating with analytics functions and orchestration systems, facilitating the execution of recommended actions and the validation of their outcomes.
Within this framework, the operational domains function as an interconnected stack rather than isolated silos.



Insights generated in one domain, such as the validation of a customer-impacting issue, can trigger deeper analysis in another domain, such as root cause identification, which in turn informs optimization or remediation recommendations. These recommendations are then executed through the AIOps layer, which manages workflow orchestration and ensures that actions are applied consistently and tracked for effectiveness.

This closed-loop, cross-domain coordination enables continuous feedback between detection, analysis, decision-making, and execution, forming the foundation for autonomous network operations. By maintaining alignment among analytics, AI-driven insights, and orchestration mechanisms, AIOps ensures that the system can not only identify and understand issues but also act on them and verify outcomes in a controlled, iterative manner.

> Closed-loop, cross-domain coordination is the foundation for autonomous network operations

RADCOM Neura: AI Agents for Autonomous Service Assurance

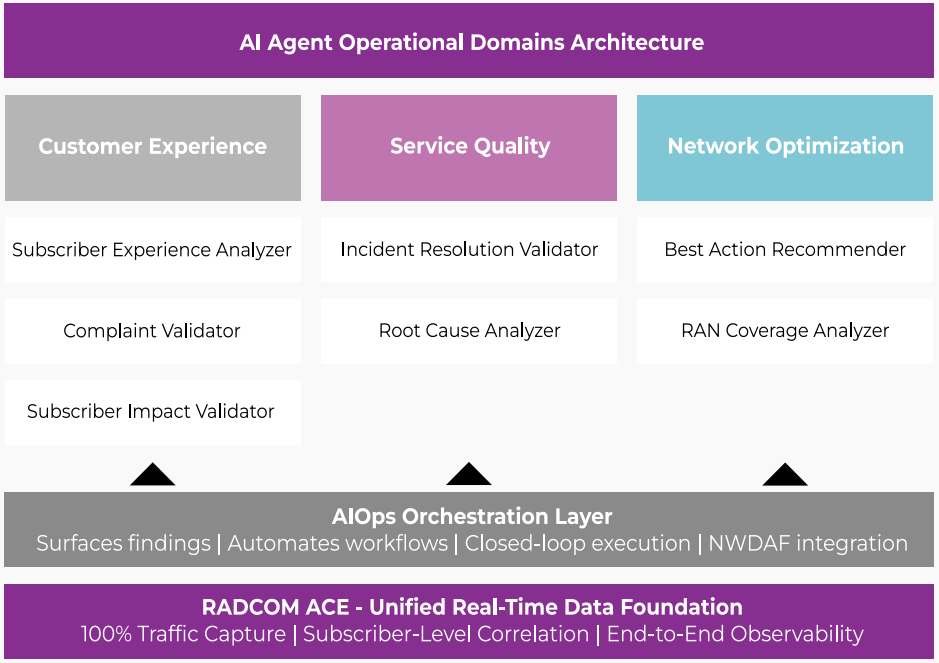


Figure 5: RADCOM Neura agent architecture - four domains on a shared RADCOM ACE data foundation

> RADCOM Neura is powered by trusted, unified assurance data with a complete real-time view of the network and customer experience

RADCOM Neura is an AI agent suite built natively within RADCOM ACE, RADCOM's cloud-native service assurance platform. It brings the agent framework described in the preceding sections from concept to production, deployed today within live network environments.

Within this framework, agentic AI is powered by trusted, unified assurance data, ensuring that every decision is based on a complete, real-time view of network performance and customer experience. This data foundation enables AI agents to move beyond analysis into action, forming the backbone of closed-loop automation.

The agents referenced throughout this paper are not theoretical constructs; they are active components of RADCOM Neura, currently deployed with Tier-1 operators running 5G Standalone (SA) networks. These deployments demonstrate the application of agentic AI in real-world, large-scale telecom environments, supporting the transition to autonomous operations across customer experience, service quality, network optimization, and AIOps domains.

RADCOM Neura's Data Foundation

RADCOM Neura operates within RADCOM ACE's cloud-native, Kubernetes-orchestrated architecture. **The platform is inherently elastic, scaling dynamically to accommodate fluctuating traffic volumes without manual intervention,** and is designed to integrate within multi-vendor RAN and core environments. NWDAF integration is natively supported, enabling the platform to both consume analytics from the 5G Core Network Data Analytics Function and publish insights back into it. This bidirectional interaction supports closed-loop coordination with orchestration systems, aligning with the requirements of Level 4 autonomous network operations.

RADCOM's Intelligent Copilot (NetTalk™)

RADCOM NetTalk™ is a GenAI-powered conversational interface that provides a human-in-the-loop access layer for operators, utilizing RADCOM Neura and RADCOM ACE. It enables users to query the underlying data behind AI-generated recommendations, validate findings prior to executing automated actions, and investigate anomalies that fall outside the current scope of agent-driven analysis. This ensures transparency and control across operations, with every agent action explainable and every recommendation traceable back to the underlying subscriber-level data that informed it.

Conclusion: Single Source of Truth

Autonomous networks rely on more than just advanced AI. They need a unified, real-time data foundation that allows AI agents to operate with full end-to-end visibility and context. As 5G drives unprecedented scale, complexity, and dynamism across the RAN, core, and service layers, the ability to unify, contextualize, and operationalize data becomes the determining factor in achieving autonomy.

RADCOM ACE is the single source of truth that AI agents need to operate autonomously. By combining continuous, subscriber-level observability with AI agents capable of cross-domain reasoning and closed-loop execution, operators can move from reactive operations to fully autonomous networks. In this approach, networks are not only monitored and analyzed but also continuously optimized in real time, with issues identified and resolved before they affect the customer experience.

For more information on RADCOM Neura, or to schedule a demo, please contact us at info@radcom.com

We Can Help Today

For further information: visit www.radcom.com or contact us for a demo.