

FROM NOISE TO INTELLIGENCE

By Rick Fulwiler, NETSCOUT.

How Data Curation Unlocks AI in Mobile Networks.

Artificial Intelligence (AI) holds real promise for mobile network operators — from automated fault detection to predictive service assurance and real-time fraud prevention. But there is a fundamental problem that many organizations underestimate: AI agents are only as good as the data you give them, and most network data, in its raw form, is effectively unusable by AI.

While generative AI agents and large language models are powerful, they are not designed to directly ingest raw 3GPP packet data generated by the radio access network (RAN) or mobile packet core. In mobile environments, meaningful AI outcomes depend on curated, structured intelligence generated upstream through domain-specific analytics.

Modern mobile networks operate at extreme scale, with a 10 million Subscriber network generating close to a million transactions per second and more than a petabyte of data per day. At this volume and velocity, raw packet-level telemetry is not only impractical for AI systems to process—it is counterproductive. Even if an AI model could technically ingest such data, the result would not be intelligence, but noise. Excessive, non-curated data overwhelms models, increases the risk of hallucinations, and leads to unreliable conclusions. More data does not automatically translate into better outcomes.

The Scale Problem: Why Raw Packets Break AI

Modern mobile networks generate staggering volumes of data. Across the radio access network (RAN) and mobile packet core, a CSP with 10 million subscribers can produce close to one million transactions per second and more than a petabyte of data per day. That is not a data storage challenge — it is a data comprehension challenge.

Feeding raw 3GPP packet data directly into an AI model does not produce insight. It produces noise. The model has no way to distinguish a meaningful event from background traffic, no understanding of session state, and no awareness of the relationships between packets flowing across different network domains. The result is an increase in hallucinations, false positives, and unreliable outputs.

More data does not automatically mean better AI. In fact, the opposite is often true: non-curated data actively degrades model performance.

What Data Curation Actually Means

Data curation is the process of transforming raw packet-level traffic into structured, contextualized information that AI systems can reason about. This is not simply filtering or compression — it is about adding contextual intelligence to the data itself.

Effective curation requires the network data platform to track every subscriber across their full session lifecycle, not just capture isolated packets. That means correlating attributes such as:

- Location — where the subscriber is at each point in the session
- Timing — when events occurred and in what sequence
- Session flow — how traffic traversed the network end-to-end
- Network domain — whether activity originated in the RAN, core, or transport layer
- KPIs and metrics — quantitative indicators that characterize normal and abnormal behavior

This enrichment process creates what can be described as an ontology — a structured representation of network reality that gives AI models the situational awareness they need to draw accurate conclusions. Without it, even the most capable AI agent is operating blind.

MCP as the Integration Layer for Network AI

Once network data has been curated and enriched, the next challenge is getting it to AI agents in a consistent, scalable, and maintainable way. This is precisely the problem that Model Context Protocol (MCP) is designed to solve.

In a mobile network architecture, the MCP server acts as a structured gateway between the data curation layer and any AI agent or large language model (LLM) that needs to query network state. Rather than each AI application building its own custom connector to proprietary data stores, MCP provides a standardized interface that any compliant AI agent can use.

Why Traditional Data Sources Lie by Omission

Before discussing what AI can do with curated packet data, it is worth addressing a common misconception: that network operators already have sufficient visibility through existing data sources. They do not. Alarms, logs, and Network Equipment Manufacturer (NEM) event streams are useful operational tools, but they share a fundamental flaw — they only report what they were explicitly designed to report. Everything else is silence.

This is not a configuration problem that can be solved with more tuning. It is structural. An alarm fires when a pre-defined threshold is crossed. A NEM event is generated when the equipment vendor anticipated a condition worth logging. A log entry exists because someone, at some point, decided that event was worth recording. All three sources are shaped by assumptions made in the past, about conditions that were known at the time. Novel failure modes, slow-burn degradations,

and multi-domain interactions that fall between vendor-defined categories simply do not appear. The network looks healthy. No alarms are firing. No events are queued. The subscriber experience is silently deteriorating.

Consider the specific blind spots each source carries:

- **Alarms** are threshold-driven and binary. A cell site either crosses a congestion threshold, or it does not. The gradual erosion of session quality that precedes a threshold breach — the kind of sustained, sub-threshold degradation that directly impacts user experience — produces no alarm and leaves no trace in traditional monitoring systems.
- **Logs** are domain-scoped and retrospective. A core network log records what happened within that element, but it has no visibility into what occurred in the RAN or the transport layer during the same session. Correlating across domains requires manual effort, specialist knowledge, and often fails entirely when timestamps are misaligned or log formats differ between vendors. Logs also only capture what was logged — they are not a complete record of network activity.
- **NEM event streams** reflect the vendor's interpretation of network state, not an objective record of it. Each manufacturer defines its own event taxonomy, severity levels, and reporting cadence. In a multi-vendor network — which describes virtually every operator at scale — these streams are heterogeneous, sometimes contradictory, and always incomplete from the perspective of end-to-end session behavior. An event that falls outside a vendor's defined taxonomy is not reported. It never existed.

The consequence for AI is severe. A model trained on alarms, logs, and NEM events will reason about the network as those sources present it, not as it truly exists. It will inherit every blind spot, every threshold boundary, every vendor-imposed category. When an anomaly occurs that does not fit the established taxonomy, the AI will either miss it entirely or misclassify it. The garbage-in, garbage-out problem is not about data volume. It is about data completeness.

Packet data is different. Every transaction traversing the network is captured, regardless of whether any element raised an alarm, generated an event, or wrote a log entry. Packet data does not require a threshold to be breached. It does not depend on a vendor deciding that something was worth reporting. It is a continuous, unfiltered record of what happened on the network — including the failures, the degradations, and the anomalies that every other source missed. That is what makes it the ground truth.

Service Assurance and Security Use Cases

When curation and MCP are in place, AI becomes genuinely useful across a broad range of operational scenarios.

Service Assurance

An AI agent connected via MCP to a curated session data store can analyze network behavior end-to-end — understanding how individual sessions perform over time, where degradation begins, and whether issues are isolated or systemic. Instead of reacting to generic threshold alerts, operators can ask the AI agent natural-language questions: “Why are VoNR sessions degrading in the North Dallas cell ID XYZ ?” and receive a contextualized, evidence-based answer.

Security Operations

The same curated datasets that power service assurance also enable a range of security use cases. These include:

- Detecting rogue base stations through anomalous signaling patterns
- Identifying packet flood attacks by correlating volume spikes with session context
- Flagging SIM swap fraud by detecting session location inconsistencies
- Identifying location misuse through cross-domain session correlation

In each case, success depends on the AI agent having access to relationship-aware, contextualized data — not just raw traffic counts. MCP provides the structured access layer that makes this possible.

Real-World Example: Mobile Banking with MCP-Backed Network Context

A practical illustration of this architecture in action: a third-party mobile banking application needed to support real-time international money transfers between subscribers in different countries. To do this reliably and securely, the application needed to verify subscriber location and session integrity at the time of each transaction.

Initial approaches — querying traditional network databases (HLR/VLR) directly or ingesting Call Detailed Records (CDR) or data packet streams — failed. The databases could not handle real-time query volumes, and packet data streams overwhelmed the application.

The solution was a data curation layer exposed via an MCP server. The banking application’s AI logic could query subscriber session state — location, session continuity, network domain — through a clean, low-latency MCP interface. The banking application’s AI only saw what it needed: a structured, queryable representation of subscriber context. The result was transaction processing at second-level resolution, with no excessive overhead and a measurably better user experience.

Conclusion: Intelligence Is About Understanding, Not Volume

The path to AI-driven mobile network operations is not paved with more data — it is paved with better-understood data. Raw packets, at network scale, are beyond the reach of any AI agent working directly. Curation, correlation, and contextual intelligence enrichment are not pre-processing steps you can defer; they are the foundation on which everything else depends.

Model Context Protocol adds the final piece: a standardized, scalable, and economically efficient interface that connects curated network intelligence to AI agents — whether for service assurance, security operations, or third-party applications. As AI adoption in mobile networks accelerates, the operators who invest in curation infrastructure and MCP integration will be the ones who realize the technology’s full potential.

In mobile networks, the question is never how much data you can collect. It is how well you can make an AI understand it.

LEARN MORE

For more information about NETSCOUT solutions visit:

www.netscout.com



Corporate Headquarters
 NETSCOUT Systems, Inc.
 Westford, MA 01886-4105
 Phone: +1 978-614-4000
www.netscout.com

Sales Information
 Toll Free US: 800-309-4804
 (International numbers below)

Product Support
 Toll Free US: 888-357-7667
 (International numbers below)

NETSCOUT offers sales, support, and services in over 32 countries. Global addresses, and international numbers are listed on the NETSCOUT website at: www.netscout.com/company/contact-us