Autonomous intelligence in RAN

ERICSSON

Contents

1 Executive summary

Mobile networks are entering an era in which competitive advantage based merely on spectrum holdings, raw throughput or hardware refresh cycles is insufficient; it is the intelligence in the network and its ability to operate autonomously that will make the difference. Operators today are navigating a structural shift: traffic patterns are becoming increasingly volatile, new service categories are emerging without notice and open, multi-vendor architectures are bringing both opportunity and operational exposure. At the same time, the complexity of tuning, optimizing and assuring radio access network (RAN) performance continues to escalate. A single Next-Generation Node B (gNB) exposes thousands of tunable parameters, and the interactions across the transport, RAN and core networks create an operational environment too large for manual engineering or traditional automation to manage reliably.

This paper addresses the central question facing operators:

If we want to modernize our networks toward zero-touch operations, differentiated connectivity and sustained OPEX efficiency, where should we invest first—and why?

The answer is clear:

Autonomy begins in the RAN.

Intent frameworks, orchestration platforms and rApps provide essential translation and governance, but the decisive point where intent becomes real—where latency, spectrum efficiency, energy consumption, mobility performance and Service Level Agreement (SLA) adherence are actually delivered—is inside the RAN's execution loops. It is the gNB that ultimately determines whether a business or service objective is fulfilled.

For operators, this reality translates into a shift in modernization strategy. Investments must prioritize the layers with the highest structural impact: enhancing the RAN as an intelligent execution environment, building robust observability and data pipelines, and adopting an automation architecture – such as Ericsson Intelligent RAN Automation – that allows intent and distributed intelligence to scale predictably across domains.

1.1 Key takeaways for operators

- **Intent realization happens in the RAN**
  Orchestration can interpret and translate high-level goals, but only the gNB can execute them with the precision required for service differentiation and enterprise-grade SLAs.
- **Intelligent RAN automation enables a manageable low-OPEX level**
  By establishing consistent intent flows, model life cycles and the placement of intelligence, Ericsson Intelligent RAN Automation allows operators to scale slicing and differentiated connectivity without driving OPEX upward.
- **Intelligence is central to reaching higher autonomy levels**
  While rApps strengthen orchestration and life-cycle governance, the greatest business value emerges when intelligence is present inside the RAN, where live outcomes are shaped.
- **Investment in RAN intelligence yields the highest return on investment**
  Operators realize tangible benefits – improved performance, predictable SLAs, energy savings, fewer incidents and reduced integration overhead – when the RAN becomes a real-time decisioning environment.
- **Autonomy is an operating model, not an overlay**
  Achieving zero-touch operations requires coordinated evolution of automation, observability, data governance and organizational processes. Tools alone are insufficient without coherent architecture.

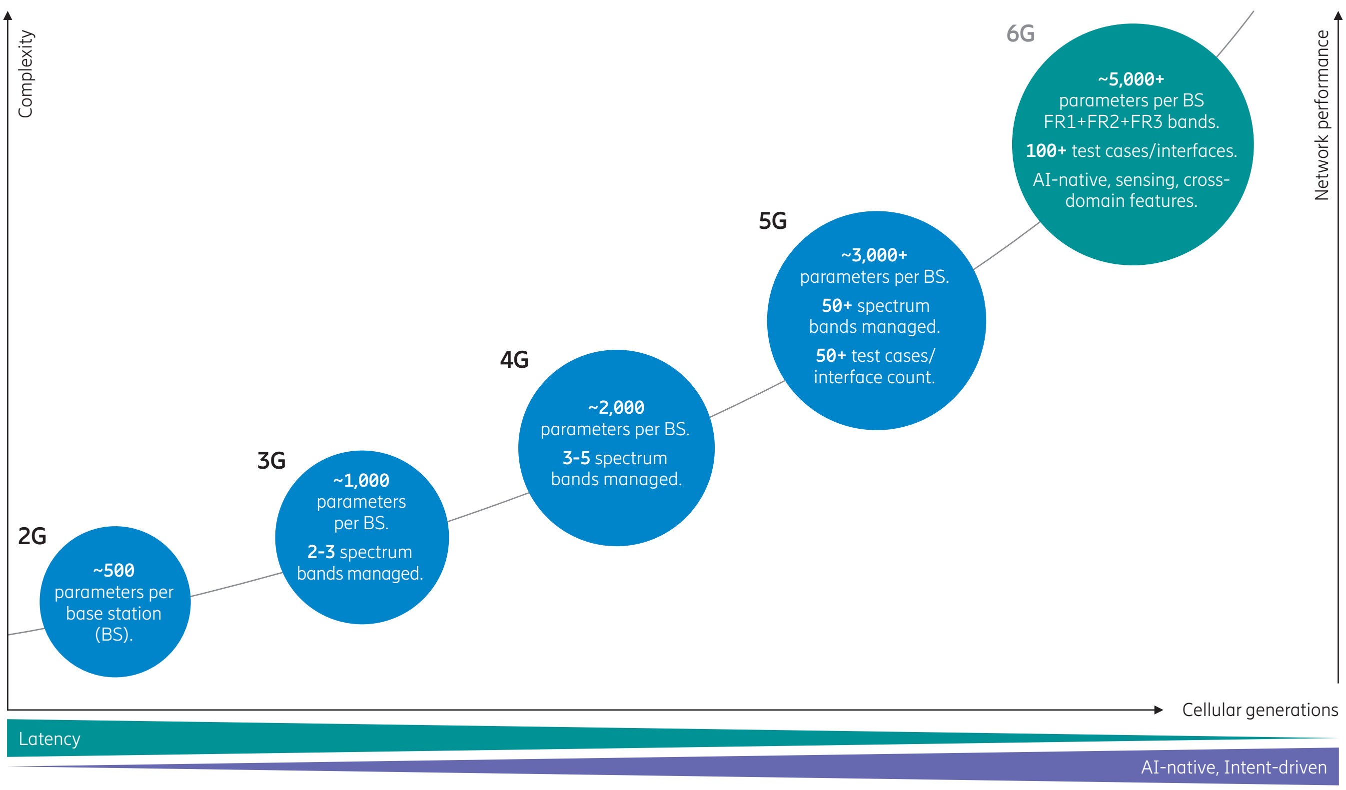
This positioning paper outlines the challenges, opportunities and modernization pathways for operators adopting intent-driven, intelligence-enabled RAN automation. It presents a clear roadmap anchored in Ericsson Intelligent RAN Automation to empower operators with the architectural foundation necessary for the next decade of autonomous networks – networks that can assure performance, minimize OPEX, support differentiated connectivity and adapt confidently to the demands of an increasingly dynamic digital world.

2 Introduction: Complexity, inflection point and convergence

Mobile networks have reached a point where complexity is no longer a by-product of growth; it is the defining constraint that operators must overcome. The once-linear evolution of radio networks has given way to an environment where the number of tunable features, interdependent layers and service expectations is growing faster than traditional tools or organizational models can support. Even a single Next-Generation Node B xNB exposes thousands of parameters, and the interactions between the transport, radio access and core networks now shape performance in ways that are difficult to anticipate or control manually.

The figure below illustrates this transformation. As networks scale geographically and in feature richness, the exponential growth of parameters and cross-domain interactions is overwhelming classical engineering and rule-based automation. Each new feature adds both capability and operational burden. Each new band or spectrum layer introduces new coordination needs. Every additional

Figure: The complexity and scale of modern network performance management



service, whether it relates to extended reality, industrial control, differentiated connectivity or mission-critical slices, brings new performance targets that cannot be managed through static rules or manual workflows.

This is not simply a complexity problem; it is a decision-scaling problem. Operators must continuously align network behavior with business outcomes under conditions that evolve minute by minute. Traditional optimization tools can only automate tasks. They cannot interpret operator goals, reason about conflicting conditions or adjust behavior coherently at the pace required.

At the same time, operators are dealing with increasingly heterogeneous environments: multi-vendor footprints, combinations of 4G, 5G Non-Standalone (NSA) and 5G Standalone (SA) modes, distributed cloud deployments and a growing set of external tools built on analytics or radio access network (RAN) data extraction. These tools provide value, but they risk fragmenting operations by operating on different data, different timescales and different control assumptions. Therefore, it is important to identify what needs to be aligned in order not to drive additional complexity when that is the core problem.

This is why operators today face an inflection point. Modernization is no longer only about spectrum strategy, coverage or capacity; it is about establishing an operating model where the network can understand what outcome is required, adapt intelligently to achieve it and verify continuously that the objective is being delivered. This is the foundation of Intelligent RAN Automation.

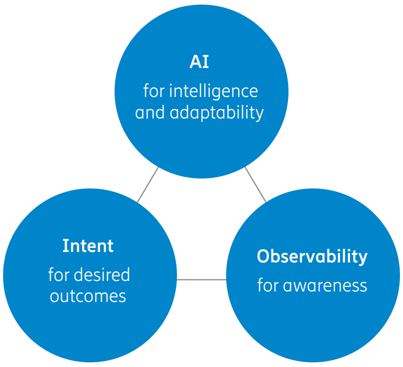
The RAN becomes central in this model because it is where customer experience, reliability, energy consumption and slice performance are determined in real time. No amount of orchestration or high-level control can compensate for the lack of intelligence or adaptability inside the node itself. To move toward zero-touch operations, differentiated connectivity and sustainable OPEX, operators must redesign how decisions are made, where intelligence resides, and how outcomes are expressed.

- Intent provides the framework to express the desired results.
- Artificial intelligence (AI) provides the ability to evaluate complex trade-offs across many interacting parameters, using data volumes and timescales that exceed what rule-based systems or manual processes can handle.
- Observability provides the data and feedback required to validate outcomes, diagnose deviations and refine AI-driven decisions over time.

Together, they enable a RAN capable not only of automation, but of autonomy.

3 Intent, AI and observability as the foundation of Intelligent RAN Automation

Figure: The combination of intent, AI, and observability forms the core foundation enabling autonomous RAN operation



As operators scale 5G and transition toward 5G Advanced, many are still operating through workflows and tools inherited from past generations. Self-organizing networks (SON) functions, operations support systems scripts and vendor-specific optimizers continue to be effective for periodic tuning, but they were designed for networks that changed slowly and behaved predictably. Today's networks no longer fit that model. Traffic patterns shift unpredictably. Spectrum layers interact in complex and non-linear ways. Mixed deployments of 4G, 5G NSA and 5G SA add further configurational complexity. In multi-vendor footprints, the coexistence of different optimization tools and external automation engines leads to fragmented decision-making.

These tools are not inadequate. They simply reflect the constraints of the era in which they were created. They automate tasks rather than outcomes, operate on partial or coarse visibility and lack any architectural way to reconcile their decisions across features, cells and domains. They cannot represent operator objectives, and they cannot influence the near-real-time behavior inside the RAN node where performance is actually realized. As networks become more dynamic and outcome-dependent, these tools alone cannot sustain operational efficiency or the performance guarantees that premium services and differentiated connectivity require.

This transition demands a new foundation built on the interplay of intent, AI and observability. These three capabilities shape the core of Ericsson Intelligent RAN Automation.

3.1 Intent as outcome-driven control

In Ericsson Intelligent RAN Automation, intent becomes the central mechanism that connects operator ambitions at the service layer with the real-time behavior of the RAN at the resource layer. Intent begins its life in the orchestration and assurance layer, where operators express what the network must deliver for a specific service, cluster, enterprise tenant or slice. These objectives can range from maintaining latency and throughput targets, to assuring indoor coverage in a venue, to meeting energy budgets with quality preserved. At this layer, intent is not yet technical. It describes the outcome, the business value or experience that the operator wants the network to deliver.

Once articulated, intent flows downward through the architecture. The Ericsson Service Orchestration and Assurance system evaluates whether the requested connectivity outcome is feasible and available. This feasibility assessment relies on historical performance, topology, current load, feature capabilities and available resources encoded through the Ericsson Intelligent Controller. Only when the orchestration layer determines that an outcome can be achieved does the intent propagate into the RAN in the form of measurable objectives and policy constraints.

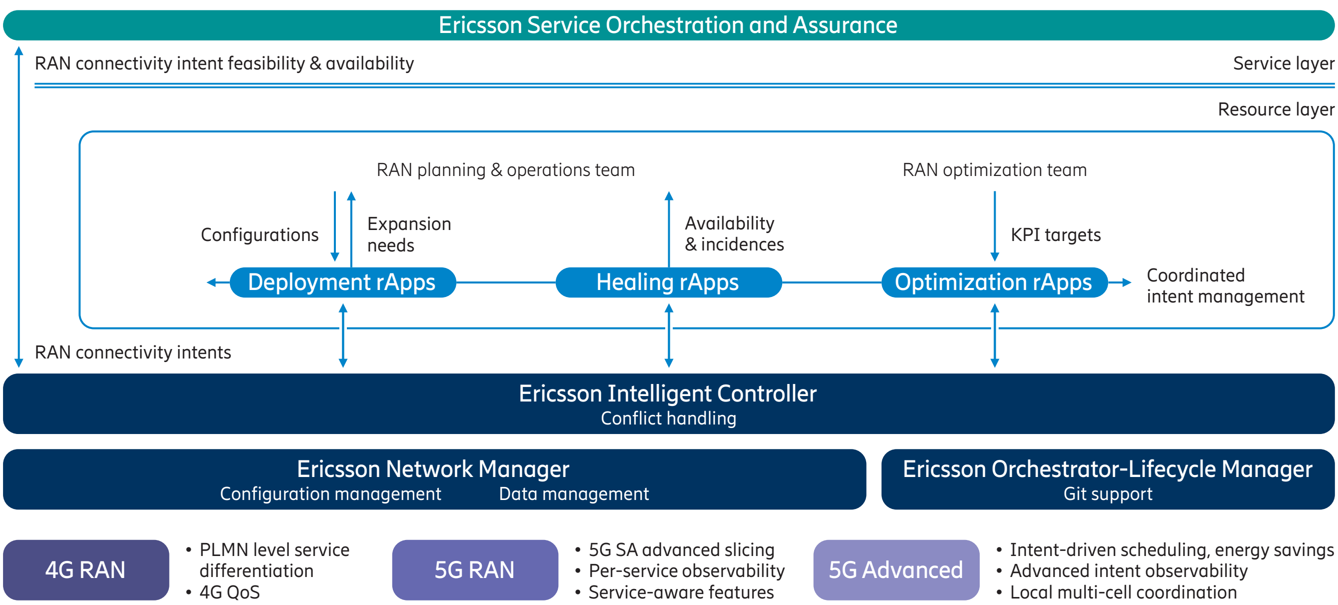
Inside the RAN is where intent becomes real work. The Ericsson Intelligent Controller distributes RAN connectivity intents to the appropriate rApps based on their role in deployment, healing and optimization. Deployment rApps ensure that new nodes or new configurations are provisioned in a way that aligns with the intent. Healing rApps monitor availability and incidents and act to maintain alignment when degradation appears. Optimization rApps translate high-level objectives into feature-level targets and policy constraints, such as KPI guardrails, resource allocation preferences, mobility behavior and interference trade-offs. These actions are coordinated across cells and features to maintain a consistent interpretation of intent, even when conditions on the ground shift.

For this process to function, each rApp must understand both the intent and its local context. The Ericsson Intelligent Controller provides the data, policy information and conflict handling needed to ensure that rApps do not satisfy opposing objectives that lead to policy conflicts. This makes the intent actionable, traceable and verifiable. It also ensures that the operator's objective remains intact as decisions move closer to the radio.

Finally, the RAN node itself becomes the point of execution. By the time intent reaches the node, it has been translated into a set of outcome priorities, KPI targets and feature-level constraints that express what matters most under the current conditions. Distributed intelligence inside the gNB interprets these priorities and targets in real time and applies them to local RAN behavior. Here, intent maps directly to how the scheduler balances competing demands, how mobility decisions are made, how energy-saving logic is applied and how other RAN features resolve trade-offs. This is the layer where latency is met or missed, where users experience continuity or degradation and where spectrum is utilized effectively or inefficiently. As intent is expressed through priorities and measurable targets rather than fixed actions, the node can adapt under varying load, interference or mobility patterns.

Through this flow, intent moves from a declarative statement into an executable control mechanism. Operators can express outcomes at the top of the system, while the RAN interprets and fulfills those outcomes in real time. Intent becomes the connective tissue between planning and execution, centralized reasoning and distributed adaptation, and business objectives and RAN behavior. It reduces fragmentation, minimizes manual intervention and enables the network to respond continuously to operator needs in a way that static automation frameworks cannot.

Figure: Intelligent RAN architecture for autonomous networks



3.2 Why intent requires AI in both the node and the automation layer

Intent has no operational value unless the network can adapt its behavior as conditions change. Modern RAN environments evolve too quickly for static rules or periodic optimization loops. Traffic surges appear within minutes. Interference landscapes shift as users move between cells. Slice demands vary with enterprise behaviors or venue usage. Radio features interact dynamically, often in ways that are difficult to predict through deterministic logic.

AI is therefore essential. It allows the network to interpret intent continuously, not only at the moment it is issued. AI in the RAN node provides the immediate adaptation required for real-time experience, for example through intelligent link adaptation, interference handling, mobility robustness and anomaly mitigation. These decisions must happen close to the radio, where milliseconds matter and context cannot be lost through centralized processing.

At the same time, centralized AI through Ericsson Intelligent Automation Platform and rApps provides the broader coordination that individual nodes cannot achieve on their own. This includes cross-cell optimization, energy savings guided by KPI guardrails, feasibility checks for new intents and the detection of systemic patterns that require region-wide action.

The architecture relies on both forms of AI. Distributed AI in the RAN aligns real-time behavior with intent. Centralized AI ensures that decisions remain coherent across clusters and aligned with the operator's larger objectives. Without this combination, intent cannot translate into predictable performance.

3.3 Observability as the foundation for trustworthy automation

Intent and AI are powerful, but without observability the system cannot ensure that the outcome is being met or even understood. A network cannot be optimized, adapted or trusted unless it understands its own state and how it responds to actions. Observability is therefore the starting point for intelligent RAN automation, not a supporting capability added after the fact.

Traditional observability was designed for human-driven operations. It relies on periodic counters, static KPIs and offline analysis to explain what happened after the fact. While these mechanisms remain useful, they do not scale into an autonomous, intent-driven network. As RAN behavior becomes more dynamic and decisions move closer to real time, observability itself must evolve.

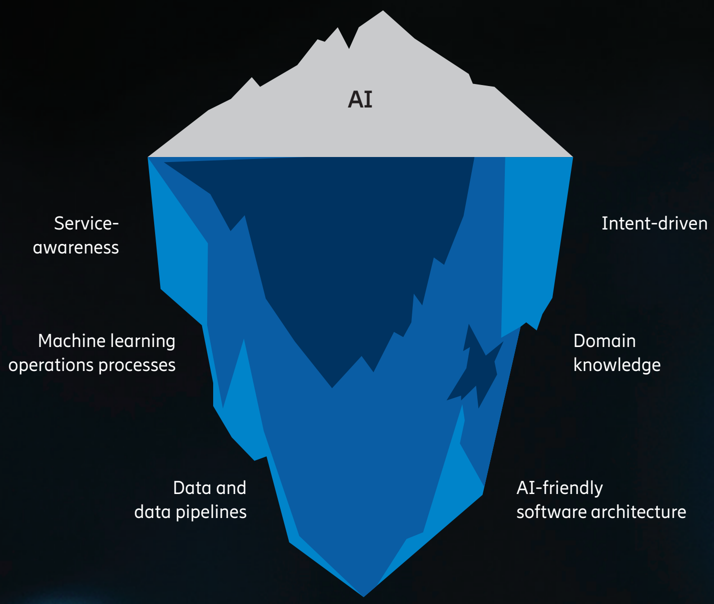
In an AI-enabled RAN, observability needs to be real-time, stream-based and adaptable. It must capture how the network behaves moment by moment and how it reacts to changes in load, configuration, interference and mobility. Just as importantly, what needs to be observed cannot be a fixed entity. An energy objective, a slice-level latency requirement or a mobility robustness target each requires different signals and different levels of granularity. Observability must therefore adapt to the intent being pursued, rather than exposing the same static set of measurements in all situations.

This shift changes the role of observability in the architecture. It is no longer a reporting layer, but an active participant in closed-loop control. It explains why a cell is stressed, why a slice is drifting from its objective or why energy consumption is rising unexpectedly. It provides early warning signals that allow corrective action before users experience degradation. It also supplies the feedback needed for AI-driven decisions to improve over time.

Digital twins extend this capability by adding prediction to observation. They allow operators to evaluate the impact of configuration changes, feature activations or new intents before applying them to the live network. Feasibility checks for intent rely on these predictive models, reducing risk and enabling the network to evolve safely toward higher levels of automation.

In autonomous networks, observability is not optional; it is the foundation that allows intent and AI to operate together in a trustworthy, scalable and autonomous manner.

Figure: Foundation needs to be in place to unleash AI potential



3.4 How intent, AI and observability work together in the Ericsson architecture

The architecture previously illustrated shows how these three elements meet inside the Ericsson Intelligent Controller and the RAN node. The service orchestration and assurance layer expresses intent, verifies feasibility and communicates availability. The Ericsson Intelligent Controller translates the operator's intent into actionable targets, distributes them to deployment, healing and optimization rApps, and reconciles decisions across domains.

Then, rApps refine these targets based on observed conditions. They monitor availability, detect deviations from expected behavior and adjust configurations in ways that align with the intent. Each rApp operates with a clear understanding of both policy and context, supported by the data exposed through the controller and the observability layer. Conflict handling ensures that the decisions of one rApp do not undermine another.

The RAN node executes the final step. It receives the KPI targets and policy constraints derived from the intent, evaluates local conditions and adapts behavior in real time. This ensures that the intent is not only understood but enacted through scheduling, interference management, mobility and energy logic.

This architecture transforms automation from a set of independent optimizers into an integrated, intent-driven system. It aligns planning with execution, central reasoning with radio-level adaptation and business expectations with measurable outcomes.

3.5 How network behavior changes when intent, AI and observability are present

To illustrate this shift, consider two examples that operators commonly face.

In energy optimization, networks that rely on static thresholds or seasonal parameters can only tune behavior infrequently. With intent-driven automation, the operator expresses the desired energy outcome, and the system observes traffic patterns in real time, predicts where energy savings are possible, and adjusts MIMO (multiple-input, multiple-output) layers or sleep states without compromising user experience. This leads to continuous and adaptive energy behavior rather than periodic tuning.

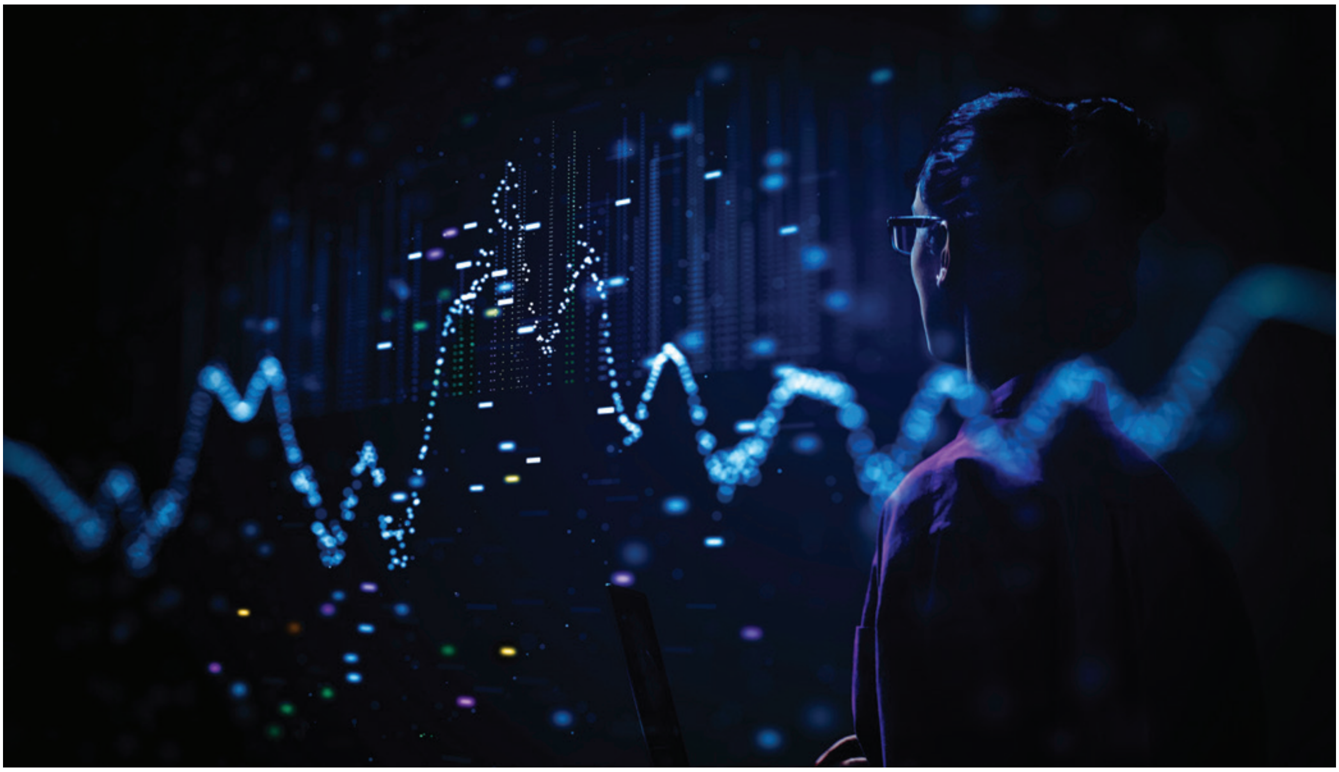
In slice assurance, 5G NSA deployments provide base-level QoS but cannot guarantee outcomes. In 5G SA areas with rich observability and AI-driven coordination, slice behavior becomes intent-shaped. The operator defines deterministic latency or throughput requirements. The system monitors slice performance, identifies emerging risks, and adapts scheduler priorities, mobility anchors or resource allocation to keep the slice aligned with its target.

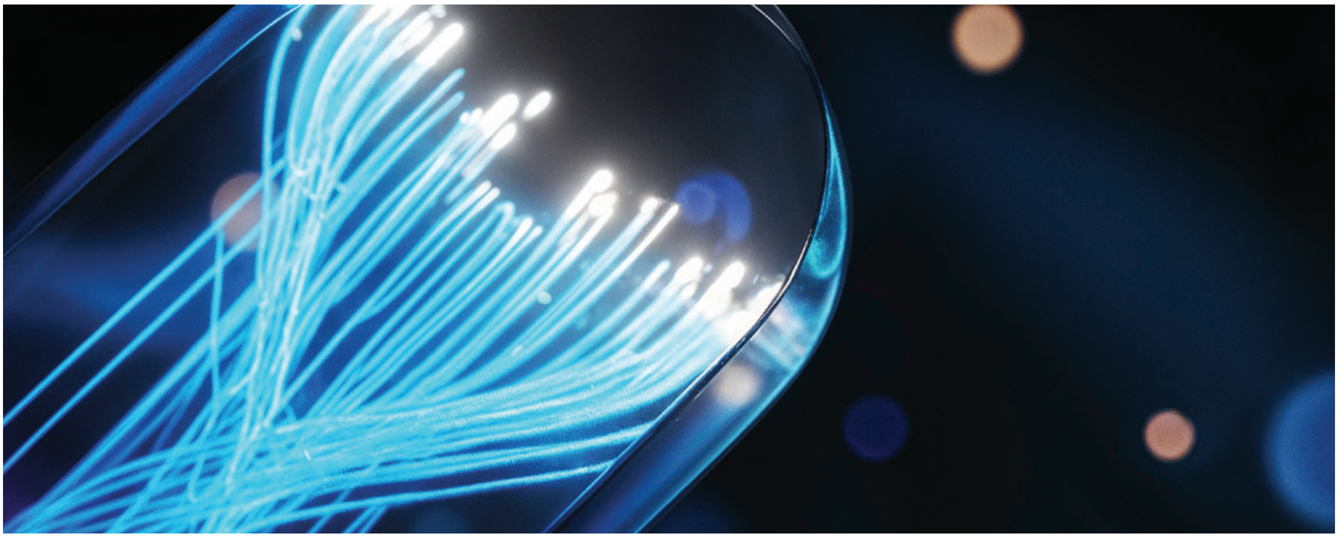
These examples demonstrate how intelligent RAN automation transforms networks into systems that understand goals, interpret context, adapt behavior and verify results. They move operators from task automation to outcome automation, where intelligence guides decisions rather than fixed workflows.

4 Operator modernization and investment pathways

Intelligent RAN automation only becomes real when operators decide where to invest and how to evolve their networks. Every operator has a different starting point, shaped by their footprint, vendor mix, spectrum portfolio, organizational maturity and the relative importance of energy, performance or revenue-generating services. The common challenge is that manual workflows, SON-era tuning and fragmented toolchains cannot scale into an outcome-driven operating model. Modernization is therefore not only a technical transformation; it is a business decision about where intelligence should reside, what capabilities create the most value, and how to balance investments between centralized automation platforms and distributed intelligence in the RAN.

Section 4 outlines how operators can evaluate their current state, understand where autonomy can create measurable value first, and identify the investment paths that allow them to evolve toward intent-driven, AI-native operations in the most practical and efficient way.



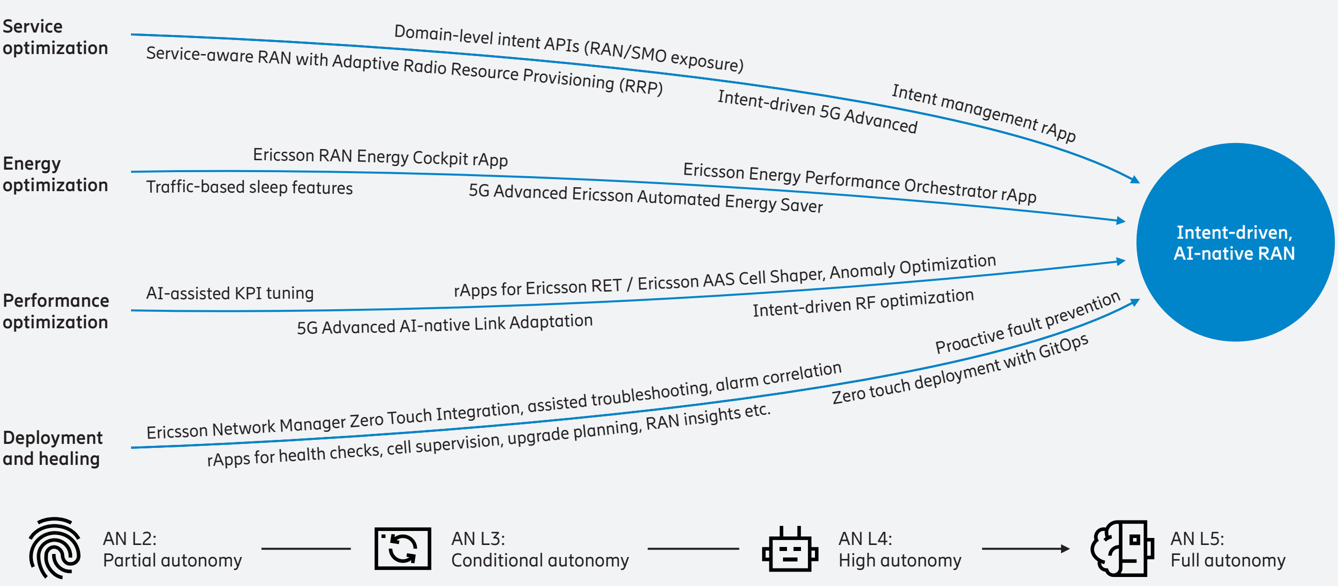


4.1 Mapping autonomy levels to Ericsson RAN features and rApps

TM Forum's (TMF's) Autonomous Network Levels provide a shared framework to describe how automation matures. What has always been conceptual becomes far more operational when mapped to the concrete capabilities in Ericsson's RAN portfolio.

The figure below shows how Ericsson has evaluated its features and rApps against the TMF levels across five operational domains: deployment, healing, performance, energy and service optimization.

Figure: RAN features and rApps for autonomous networks (ANs).



In deployment and healing, Level 2 (L2) maps to zero-touch integration and assisted troubleshooting. Level 3 (L3) corresponds to rApps that perform health checks, cell supervision, upgrade planning and insights-driven corrective actions. Level 4 (L4) includes zero-touch deployments and proactive incident prevention driven by pattern detection and policy.

In performance optimization, L2 aligns with AI-assisted KPI tuning, while L3 corresponds to AI-native link adaptation. L4 becomes possible when intent-driven radio frequency (RF) optimization aligns daily decisions with operator-defined service objectives.

In energy optimization, L2 is represented by static-traffic-based sleep features. L3 maps to the Ericsson 5G Advanced Automated Energy Saver feature and the Ericsson RAN Energy Cockpit rApp. L4 corresponds to intent-driven energy orchestration performed by the Ericsson Energy Performance Orchestrator rApp.

In service optimization, L2 includes early service-aware behavior. L3 and L4 use domain-level intent application programming interfaces (APIs) and intent-driven 5G Advanced features that allow service requirements to shape RAN behavior directly via an intent.

This mapping helps operators evaluate where they are today and where they stand to gain most. A network might be at L3 for performance, L2 for energy and L4 in a stadium or enterprise cluster. Autonomy matures unevenly. Ericsson's portfolio turns these levels into practical guidance points rather than abstract stages.

TMF level	Description	Nature of control	Operators' reality today
AN L1	Monitoring and reporting	Manual, human-driven	Wide-area 4G, rural macro, early NSA deployments
AN L2	Task automation, rules, workflows	Assisted operations	Most national networks across deployment, healing and performance domains
AN L3	Context-aware decisions with analytics and early AI	Shared human—machine control	Emerging in SA clusters, venues, enterprise zones and pilots using rApps
AN L4	Goal-oriented automation guided by policies	Coordinated machine-led action	Appears in limited high-value areas such as stadiums and private networks
AN L5	Predictive, adaptive, self-correcting behavior	Human oversight for governance only	Not present today; requires 5G Advanced and mature AI-native RAN

4.2
Typical operator starting points

Most operators fall into one of three broad starting positions. Each position implies a different modernization pathway and different initial returns.

Operators with 4G-heavy networks and SON workflows
These networks rely on distributed macros, basic SON modules and a mix of vendor-specific tools. They tend to have fragmented observability, rule-based optimization and limited automation for deployment and healing. Their modernization focus is usually on establishing observability, centralizing automation logic and relying on early rApps for deployment and performance improvements.

Operators with 5G NSA overlays
These operators have modernized radio footprints but lack the end-to-end capabilities that allow differentiated connectivity or slice-level outcomes. NSA provides throughput and capacity gains but does not provide mechanisms to assure latency, coverage differentiation or enterprise requirements. Here, the modernization focus is on data readiness, enabling AI workflows across cells, and preparing the RAN for the move to SA where intent-driven behavior becomes meaningful.

Operators with emerging 5G SA clusters
These operator networks benefit from improved service awareness, network programmability and richer data exposure compared to legacy deployments. This creates the necessary foundation to introduce AI-driven automation and move beyond rule-based optimization.

Their focus should be on enabling rApps and AI-native RAN features that can leverage network data for prediction, optimization and automation. This includes expanding data collection across cells and clusters, ensuring consistent access to performance and configuration data, and deploying rApps that support deployment automation, coverage optimization, energy management and slice-aware resource coordination.

They should also enable intent-driven operation, where service objectives such as performance targets, slice requirements or energy policies are translated into resource-level actions. rApps evaluate network-wide conditions and determine the appropriate optimizations, while distributed RAN automation and AI-native RAN features execute time-critical adaptations at the node level in alignment with the declared intent.

4.3
Centralized automation and distributed automation

Autonomy requires the right balance between centralized and distributed intelligence.

Centralized automation, enabled by rApps, operates at multi-cell and cluster scales. It performs feasibility checks, analyzes trends, reconciles conflicts and aligns decisions with operator policy. It is essential for energy management, slice alignment, deployment automation and multi-cell optimization.

Distributed automation, embedded inside the RAN node, executes decisions at real-time speeds. It manages scheduling, mobility, interference, anomaly correction and feature-rich adaptations. It is essential for user experience, SLA consistency and intent execution.

Investments in both domains are required. Centralized automation scales the network. Distributed automation ensures the network behaves correctly under real conditions.

At this stage, operators often evaluate alternative investment paths, including large-scale compute platforms, GPUs (graphics processing units) or centralized AI infrastructure. These investments can create value for analytics, customer-facing intelligence and external services. However, on their own, they do not resolve the execution challenges that dominate network operations.

Without intelligence embedded in the RAN itself, centralized compute can observe and analyze network behavior but cannot shape it in real time. Latency, mobility robustness, energy efficiency and SLA adherence are ultimately determined inside the RAN execution loops. From a business perspective, this creates an opportunity cost: capital invested in centralized intelligence delivers limited operational return if the RAN remains rule-driven or manually tuned.

Intelligent RAN automation ensures that higher-layer intelligence translates into measurable outcomes where value is actually realized, inside the RAN, rather than accumulating insight without control.

4.4 Practical modernization patterns

Operators typically combine these levers into modernization patterns that reflect their current state.

4G-heavy or SON-centric networks
These networks rely on legacy SON features, limited observability and mixed scripts or tools. In this case, the realistic path is:

- Invest first in observability and basic data readiness.
- Introduce deployment and healing rApps to reduce manual work.
- Gradually add performance and energy rApps in selected clusters where benefits are easiest to measure.

At this stage, the focus is on lifting specific domains from L2 to L3, not on immediate intent APIs.

5G NSA with scattered 5G coverage
NSA networks have modern radios but lack a clean way to assure outcomes such as deterministic latency or slice commitments. Modernization here typically:

- Uses observability and rApps to stabilize performance and reduce OPEX.
- Prepares data and processes for 5G SA clusters.
- Introduces early AI-native features in the RAN node where hardware and software allow it.

The main goal is to create SA-ready islands that can host higher levels of autonomy.

5G SA clusters and high-value zones
These are the natural candidates for intent-driven automation. They already have a core that can support slicing, differentiated QoS and advanced service constructs. The recommended sequence is:

- Strengthen observability for SA clusters, including slice-level KPIs.
- Deploy performance, healing and energy rApps and align them to cluster objectives.
- Enable AI-native features in the RAN node where available.
- Introduce intent APIs for selected high-value services, such as industrial tenants or venue connectivity.

These steps move targeted zones from L2 to L3 and in some cases to L4, while the wider footprint continues to advance more gradually.

From investment levers to a decision pathway
The combination of autonomy levels, portfolio mapping and modernization patterns can be expressed as a simple decision pathway for operators:

1. **Locate your current autonomy level per domain**
   Use the TMF levels and the portfolio mapping to classify where deployment, healing, performance, energy and service optimization are today.
2. **Identify the business-critical domains and zones**
   Decide where autonomy has the highest value: energy, enterprise SLAs, venue capacity, national performance or a mix of these.
3. **Select the first investment lever per domain**
   - If visibility is weak, invest in observability.
   - If visibility exists but decisions are manual, introduce rApps.
   - If rApps exist but performance is still fragile under stress, invest in AI-native features in the RAN.
4. **Introduce intent where the prerequisites exist**
   Only when observability, rApps and distributed intelligence are present in a zone does intent-driven operation make sense. At that point, intent APIs become the natural next step to unlock L4 behavior.

This pathway provides a structured way for operators to modernize without over committing. It respects existing investments, focuses on domains with clear value and builds toward an intent-driven, AI-native RAN over time.

A structured way for operators to modernize without over committing

5 Strategic recommendations for operators modernizing toward intelligent RAN automation

Choosing not to invest in intelligent RAN automation is itself a strategic decision. Networks can continue to operate using manual workflows, SON-era optimization and fragmented tooling, but only at increasing operational cost and decreasing predictability.

As traffic patterns become more volatile and service expectations more differentiated, the gap between network behavior and business objectives is widening. Manual processes are scaling linearly with network complexity, while the cost of misalignment is growing exponentially. Over time, this leads to higher OPEX, slower service introduction, increased incident rates and reduced confidence in meeting enterprise-grade commitments.

In this context, the risk is not stagnation but erosion. Operators that delay investment may find themselves constrained to best-effort connectivity models, unable to reliably monetize advanced services or differentiate on quality and reliability. The question for operators, therefore, is not whether automation is necessary, but how to structure investments so that autonomy becomes a scalable operating model rather than a collection of disconnected tools.

The recommendations below outline where operators should focus and what matters most when modernizing toward intent-driven, AI-native operations.

The first recommendation is to treat observability as strategic infrastructure. Rich, reliable RAN telemetry is the foundation for every level of autonomy. It strengthens operational confidence, improves the accuracy of rApps and gives distributed AI the context it needs to adapt in real time. Operators who make the fastest progress are those who invest early in consistent measurement frameworks, data quality and multi-dimensional visibility across cells and clusters.

The second recommendation is to automate operational domains before automating intent. The leap to intent-driven networks only creates value when deployment, healing, performance and energy are already stabilized through rApps and repeatable workflows. Automating these domains reduces human load, removes fragility from large-scale changes and establishes the predictable behavior on which intent relies. It shifts automation from a set of scripts into an intelligent system.

The next recommendation is to shift adaptive capabilities into the RAN node, where outcomes are determined. AI-native RAN features provide real-time responsiveness that centralized systems cannot achieve on their own. This is essential for busy-hour performance, reliability commitments and advanced 5G service categories. Investments here yield measurable improvements in spectrum efficiency, energy savings and SLA stability, while also enabling autonomy to expand beyond a few high-value zones.

The fourth recommendation is to automate processes, not only network functions. Operators cannot scale autonomy with manual rollouts, manual pre-checks or manual validation workflows. Process automation for tasks such as zero-touch upgrades, automated verification, guided root-cause analysis and model life-cycle governance removes internal friction and accelerates the deployment of new capabilities. It is process automation that turns domain-level improvements into a sustainable operating model.

The final recommendation is to introduce intent only where the prerequisites exist. Intent becomes powerful when the network can understand it, act on it and verify the outcome without human interpretation. This requires observability, rApps and distributed AI to already be in place. Once these elements mature, intent becomes the natural enabler for differentiated connectivity, enterprise SLAs and new service categories that reflect business priorities rather than network constraints.

Together, these recommendations provide operators with a modernization strategy that is practical, sequenced and measurable. They offer a way to reduce OPEX pressure, elevate service differentiation and prepare for 5G Advanced without unnecessary risk. Ericsson Intelligent RAN Automation provides the architecture and capabilities needed to execute this strategy, allowing operators to evolve at their chosen pace, while building the foundations for the next decade of mobile network operations.

Ericsson Intelligent RAN Automation is not a feature; it is the operating model of modern networks.

By aligning intent, AI, observability and process automation, operators can turn complexity into predictable outcomes.

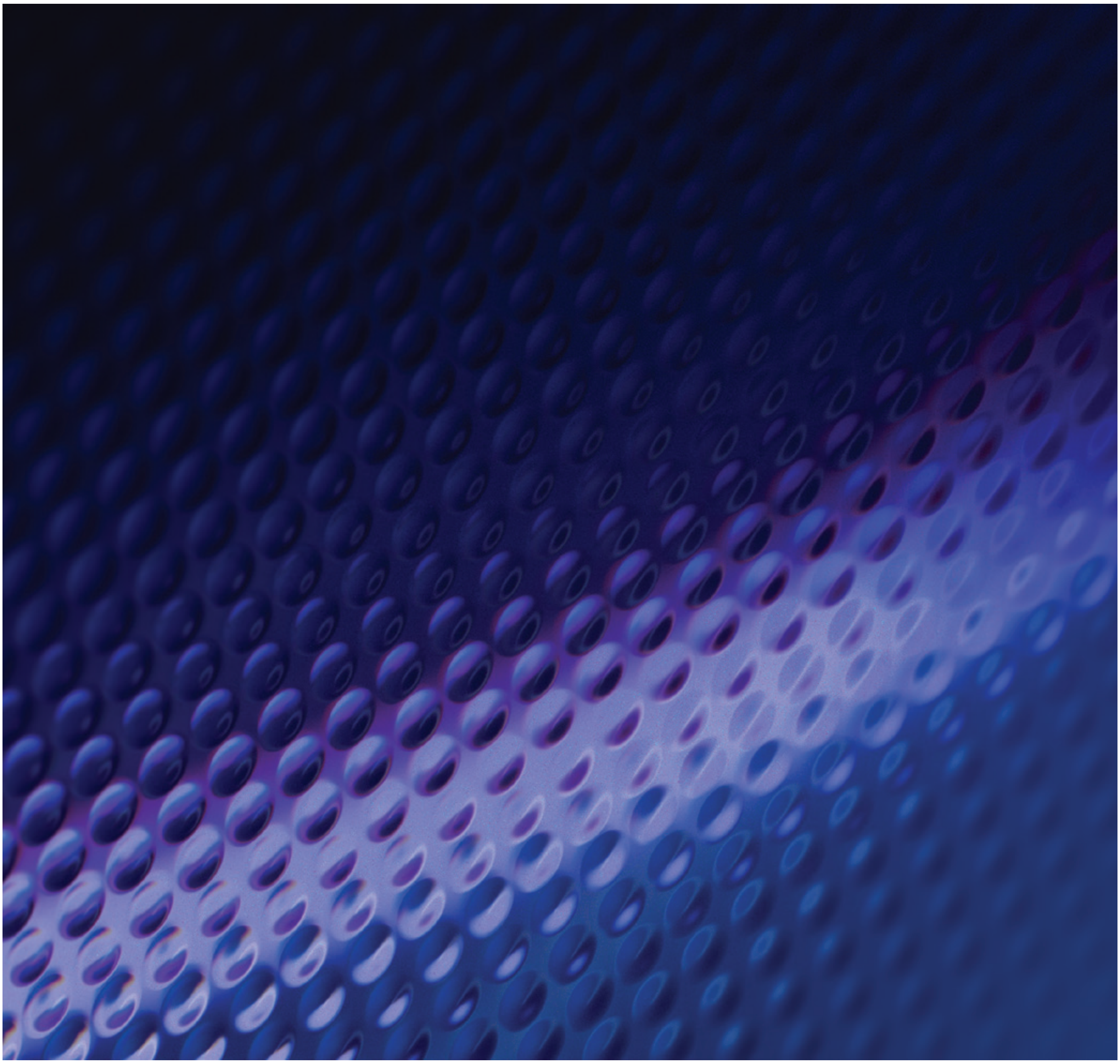
Ericsson provides the pathway, the portfolio and the architecture to make that shift real.

Authors

Christina Chaccour
Senior Strategy and Product Manager,
Business Area Networks, Ericsson

Sara Karlberg
Head of Architecture, Automation and Operations,
Business Area Networks, Ericsson

Klas Johansson
Head of Services and Security,
Business Area Networks, Ericsson



1/0341-FGD 101 495 Uen