Beyond Capacity:

Redefining RAN Strategy in the Age of AI and 5G

A White Paper by FL Omnitele

May 2026

FUSIONLAYER

Annankatu 27
00100 Helsinki
Finland

info(at)fusionlayer.com
+358 75 325 2992
www.fusionlayer.com

By FusionLayer Group, May 2026.

Any comments relating to the material contained in this document may be submitted to:

FusionLayer Inc.
Annankatu 27, FI-00100 Helsinki, Finland.
or by email to: info@fusionlayer.com

Executive summary

For the better part of the past decade, the governing logic of RAN planning was simple: mobile data traffic was growing at 40–50% annually, and investment strategy was driven overwhelmingly by capacity. That era is over. Traffic growth has moderated globally to approximately 20% per year and is expected to continue stabilizing. The relief on capital expenditure is real — but it is also a trap.

Operators that interpret slowing traffic growth as a signal to reduce strategic ambition will find themselves unprepared for three simultaneous and transformative forces:

1. The shift in focus from capacity to revenue
2. The AI-driven progress from downlink-centric to uplink-sensitive traffic patterns
3. A multi-year technology migration from legacy 2G/3G to 5G — and eventually to 6G from the 2030s onward.

This paper examines each of these forces and offers a framework for how Operators can reorient RAN planning from a cost-efficiency discipline into a revenue-generation capability, while protecting the network performance and customer trust that is the ultimate foundation of long-term competitiveness. FL Omnitele stands ready to support Mobile Network Operators (MNOs) through this transition.

Annankatu 27
00100 Helsinki
Finland

info(at)fusionlayer.com
+358 75 325 2992
www.fusionlayer.com

1. The End of the Capacity Supercycle

For years, the primary question in RAN investment was not whether to expand capacity, but how fast. Annual mobile data traffic growth rates of 40–50% created an environment in which operators were perpetually building to keep pace with demand. Network planning was essentially capacity planning. As revenue growth remained in the single digits, much of the focus on RAN investments was on cost efficiency.

That structural condition has now changed. According to recent reports, global mobile network data traffic grew 20% between Q3 2024 and Q3 2025 — a figure that, while still significant in absolute terms, represents a substantial deceleration from prior years. The same reports note this rate of growth was 'slightly above expectations,' suggesting that the underlying trajectory is one of gradual normalization rather than a temporary dip.

20%	**2.9bn**	**6.4bn**
Annual mobile data traffic growth Q3 2024 – Q3 2025	5G subscriptions expected by end of 2025	5G subscriptions forecast by 2031

This moderation has meaningful implications. Operators now face a different resource allocation challenge: how to redirect engineering talent, capital, and strategic attention from pure capacity management towards value creation. The pressure is not gone, but instead, it has shifted.

"The governing logic of RAN planning is shifting — from capacity as the answer to capability as the question."

Annankatu 27
00100 Helsinki
Finland

info(at)fusionlayer.com
+358 75 325 2992
www.fusionlayer.com

2. Monetizing AI: The New Strategic Imperative for RAN

If capacity was the lens through which the prior decade of RAN investment was understood, Artificial Intelligence (AI) is the lens through which the next decade will be framed. This is not just a technology observation, but rather a monetization argument. The question facing every operator is no longer whether AI will transform their networks, but whether their networks will be able to generate revenue from that transformation.

2.1 AI for Operational Efficiency: The Foundation Layer

The first and most immediate avenue for AI monetization is internal: using artificial intelligence to operate networks more efficiently. This includes energy optimization, autonomous fault detection, predictive maintenance, and self-optimizing network capabilities.

Also, energy efficiency deserves particular attention. As 5G networks expand and the number of AI workloads increases, energy costs are increasingly a share of Operating Expenses (OPEX). In this area, AI-driven energy management, including deep sleep modes during low-traffic periods and other energy-saving measures, offers direct cost reduction at scale. This is not a future promise but rather an existing way to optimize costs and increase Return on Investment (ROI) immediately.

2.2 AI as External Revenue: Serving New Traffic Architectures

The more strategically significant yet also more demanding dimension of AI monetization lies in serving AI-generated traffic. This is where RAN planning assumptions must be revised most thoroughly.

Annankatu 27
00100 Helsinki
Finland

info(at)fusionlayer.com
+358 75 325 2992
www.fusionlayer.com

2.2.1 THE UPLINK REVOLUTION

Mobile networks have historically been designed around downlink dominance: content streamed to devices. AI inverts this asymmetry. As research across more than 50 AI applications confirms, AI-driven traffic shows a pronounced shift toward higher uplink volumes. Large language model interactions, agentic AI systems, and particularly Physical AI — robotics, autonomous vehicles, augmented reality, and drones — all generate substantial, often latency-sensitive upstream data flows.

The research describes Physical AI as potentially requiring 'a fundamental change in how traffic is handled in the radio access network.' The core challenge of Physical AI is that it relies on low-latency uplink video for real-time control with driverless vehicles requiring remote operator assistance, service robots needing guidance in complex environments, and delivery drones requiring real-time analysis as typical examples. Unlike best-effort content delivery, these use cases do not tolerate increasing latency as traffic increases.

2.2.2 FDD VS. TDD: A STRATEGIC SPECTRUM QUESTION

The uplink opportunity raises a critical planning question about spectrum configuration. Frequency-Division Duplex (FDD) spectrum, which separates uplink and downlink into distinct frequency bands, can provide symmetric bandwidth for uplink capacity. However, it faces growing bottlenecks as AI traffic volumes surge. Time-Division Duplex (TDD) spectrum, with more bandwidth available for mid-band 5G, can offer greater flexibility in allocating capacity between uplink and downlink, making it potentially better suited to the emerging traffic mix.

This is not purely a technical question. Regulatory decisions on spectrum allocation and the licensing conditions attached to FDD and TDD bands will substantially shape how quickly operators can adapt their uplink capacity. MNOs that engage proactively with regulators on spectrum policy — framing the case for uplink-oriented flexibility — will be better positioned than those who wait for the regulatory environment to shift around them.

Annankatu 27
00100 Helsinki
Finland

info(at)fusionlayer.com
+358 75 325 2992
www.fusionlayer.com

2.2.3 AI-RAN: THE MULTI-PURPOSE COMPUTING PLATFORM

Beyond traffic handling, AI-RAN represents a more fundamental architectural evolution: transforming the RAN from a connectivity layer into a distributed computing platform. In this vision, radio sites become edge inference nodes that host AI workloads alongside traditional RAN functions.

The commercial timeline is important to calibrate. Widespread AI-RAN deployments are expected closer to 2028. Operators should treat the current period as one for strategic preparation — building the architectural understanding, vendor relationships, and regulatory positioning that will determine readiness when the window opens

01	02	03
AI for RAN	**AI on RAN**	**AI and RAN**
Using AI to optimise network operations: energy efficiency, self-healing, autonomous SON, predictive maintenance. Near-term ROI available today.	Serving AI-generated traffic with guaranteed SLAs: Physical AI, uplink-intensive applications, latency-sensitive agentic workloads. Requires network redesign from 2025-2028.	RAN as distributed edge compute platform for AI inference. New revenue streams beyond connectivity. Commercial scale expected from 2028 onward.

Annankatu 27
00100 Helsinki
Finland

info(at)fusionlayer.com
+358 75 325 2992
www.fusionlayer.com

3. Managing the technology generation transition

Alongside the AI monetization agenda, operators face the practical and operational challenge of managing a complex multi-generational technology transition. The trajectory is clear: 3G sunsetting is underway globally; 4G remains the dominant access technology by traffic volume but is declining in subscription share; 5G is scaling rapidly and approaching majority subscriber penetration in leading markets, only held back by low-cost device availability; and 6G standardization has begun, with commercial launches anticipated to accelerate in the 2030s.

3.1 The 5G Inflection Point

Studies show 5G subscriptions will reach 2.9 billion by the end of 2025, accounting for approximately one-third of all mobile subscriptions globally, with North America at 79% penetration and Western Europe at 55%. The share of traffic carried over 5G continues to rise. As a consequence, 5G is forecasted to overtake 4G measured in subscription count by the end of 2027.

Yet despite rapidly scaling 5G subscription volumes, revenue per user globally is under pressure. The volume-value gap — more users, more data, flat or declining revenue per unit — is the defining financial challenge of the 5G era.

The answer lies in using 5G capabilities — network slicing, guaranteed SLAs, private campus networks, enterprise services — to create new monetizable products, rather than treating 5G as a faster version of 4G. This reframing is precisely what shifts RAN planning from a cost discipline to a revenue capability.

3.2 The 6G Horizon: Planning with Discipline

6G standardisation through 3GPP Release 21 (IMT-2030) has begun. Commercial launches are expected in leading markets — the US, China, Japan, South Korea, the GCC — in the early 2030s, with European deployments approximately one year later due to the delayed 5G SA rollout. Global 6G subscriptions are forecast to reach 180 million by the end of 2031, though this figure excludes AI-enabled IoT devices such as autonomous vehicles and smart glasses, which could significantly accelerate uptake. The industry has learned hard lessons from 5G: the risk of heavy capital commitment to a generation before use cases and monetization models have matured.

Annankatu 27
00100 Helsinki
Finland

info(at)fusionlayer.com
+358 75 325 2992
www.fusionlayer.com

Capital Expense (CAPEX) as a share of Telecom revenue has already reduced from 26.9% in 2022 to 22.9% in 2024. Even 5G-Advanced and 6G in the late 2020s and the 2030s are expected to cause only a marginal uptick once the rollouts begin.

3.3 Legacy Sunsetting: The Trust Preservation Imperative

Perhaps the most underappreciated challenge in the technology transition is the operational risk posed by the sunsetting of legacy networks. While the strategic imperative to migrate customers onto modern infrastructure is clear, the execution risk is not trivial. Millions of customers — including enterprise IoT devices, critical infrastructure, and long-tail consumer segments — remain dependent on legacy technologies. Service degradation during migration is not a minor inconvenience; it is a trust event.

> *Customer trust in network quality is the product of years of consistent experience. It can be built incrementally but lost overnight. Operators who treat legacy sunsetting as purely a cost reduction exercise, without the discipline of protected customer experience through the transition, risk precisely the kind of trust destruction that undermines the premium positioning required for AI and 5G monetization.*

Effective legacy management requires a parallel-track approach: accelerating migration timelines where feasible while maintaining and protecting performance on legacy infrastructure. This is a planning and engineering challenge of genuine complexity that demands systematic, site-by-site analysis rather than top-down timeline mandates.

Annankatu 27
00100 Helsinki
Finland

info(at)fusionlayer.com
+358 75 325 2992
www.fusionlayer.com

4. A New Framework for RAN Planning

The three forces described in this paper — the moderation of traffic growth, the AI monetization imperative, and the technology generation transition — collectively demand a fundamental reorientation of how operators approach RAN planning. The old framework, centered on capacity and cost efficiency, remains necessary but is no longer sufficient.

We propose that leading operators adopt a framework organized around three parallel planning disciplines:

4.1 Revenue-Oriented Network Design

RAN planning must be tied explicitly to revenue objectives, not just traffic engineering targets. This means building network capabilities—guaranteed SLAs, headroom for uplink performance, network slicing readiness, and edge compute capacity—that are prerequisites for differentiated service offerings. Sites should be evaluated not only on coverage and capacity metrics, but on their ability to support premium enterprise and AI-era services.

4.2 Uplink Capacity Planning

The traditional downlink-centric planning model must be rebalanced. As AI traffic grows and Physical AI use cases emerge, uplink capacity and latency performance will become a new requirement for RAN planning. Operators should audit their spectrum portfolios for uplink flexibility, engage regulators on FDD/TDD allocation, open a dialogue with the competitors, and begin modeling uplink bottleneck scenarios — particularly in dense urban areas where Physical AI adoption will accelerate fastest.

4.3 Resilient Technology Migration

The transition from legacy to 5G and from 5G to 6G must be managed as a customer-experience risk, not only as a capital program. This requires site-level migration planning, performance monitoring through transition windows, and explicit commitments to service continuity. Trust is the operator's most valuable and most fragile asset. Network planning must treat its protection as a first-order objective alongside efficiency and coverage.

Annankatu 27
00100 Helsinki
Finland

info(at)fusionlayer.com
+358 75 325 2992
www.fusionlayer.com

5. Autonomous RAN: from Vision to Early Reality

Underpinning the AI monetization and operational efficiency agenda described in this paper is a more fundamental question of network autonomy. Specifically, to what degree can RAN operations be automated, self-optimizing, and self-healing, with minimal human intervention? According to the latest analysis, Level 4 autonomous RAN is no longer theoretical, but rather quickly becoming an operational reality for a small but growing group of MNOs.

The TM Forum Autonomous Networks framework defines a five-level scale from fully manual (Level 0) to fully autonomous (Level 5). Most operators today sit between Levels 1 and 2, where automation is domain-specific and largely rule-based. The transition toward Level 4, at which systems can reason, decide, and act with minimal human intervention, represents a fundamental shift from automating tasks to automating decisions.

5.1 The Commercial Case for Autonomy

The economic rationale for pursuing autonomy is compelling and urgent. Without automation, MNOs' OPEX risks compound at a rapid rate, further eroding already thin margins. Therefore, the efficiency case for autonomous RAN is not aspirational but rather a financial necessity.

Typical AI-driven efficiency gains demonstrated to date fall within the 10-30% range for specific RAN functions, including energy efficiency, traffic optimization, automated assurance, and interference management. The AI-RAN business case will hinge on the cost and power envelope of the required hardware. The strongest business case tends to be associated with an architecture in which automation is layered onto existing infrastructure, rather than requiring a wholesale hardware replacement.

Annankatu 27
00100 Helsinki
Finland

info(at)fusionlayer.com
+358 75 325 2992
www.fusionlayer.com

5.2 Early Proof Points: Level 4 in Live Networks

The transition from theoretical framework to production deployment is now underway. Several operators have achieved TM Forum-validated Level 4 autonomy in specific RAN domains (as of Q1 2026):

- Rakuten Mobile has demonstrated Level 4 at scale in a live RAN network, achieving 20% RAN energy savings using AI-driven closed-loop control with no impact on customer experience — validated by TM Forum.
- TDC NET and Ericsson achieved TM Forum Level 4 autonomy certification for a live RAN deployment in June 2025, focused on Ericsson's PCEM software, which reduced the energy required to transmit 1 GB of data by approximately 5% under live network conditions.
- China Mobile has reported Level 4 progress across multiple use cases, including service assurance, wireless energy optimization, and IP fault management, validated through TM Forum ANLAV assessments.
- China Telecom and China Unicom are applying targeted Level 4 automation in high-value domains such as traffic optimization and energy efficiency, with selected AI initiatives delivering double-digit efficiency improvements.

These examples demonstrate that Level 4 autonomy is achievable — but they also reveal its current character: scenario-specific and domain-limited rather than network-wide. The path to full network-wide Level 4 autonomy will be measured in years, not months.

Annankatu 27
00100 Helsinki
Finland

info(at)fusionlayer.com
+358 75 325 2992
www.fusionlayer.com

5.3. Autonomy and 6G: A Native Integration

6G will treat AI as foundational rather than additive. While 5G has required retrofitting AI and automation into an architecture not originally designed for them, 6G standardization is incorporating AI natively into areas such as the MAC layer, beam management, MIMO optimization, and scheduling. A powerful agentic layer will be essential for Level 3+ autonomous network management in 6G, handling complexity at a scale that rule-based systems cannot reach.

For operators planning their technology roadmaps today, the investments made now in autonomous RAN capabilities involving closed-loop control, AI-driven optimization, and TM Forum alignment are not merely operational improvements. Rather, they are the foundation for 6G readiness and the proving ground for the monetization models that will define the next generation.

"The transition toward Level 4 represents a fundamental shift — from automating tasks to automating decisions."

Annankatu 27
00100 Helsinki
Finland

info(at)fusionlayer.com
+358 75 325 2992
www.fusionlayer.com

6. Realize Your Goals with FL Omnitele

The strategic imperatives described in this paper are not abstract challenges. Instead, they are actionable development initiatives that require concrete analytical capability, deep RAN expertise, and the experience to translate strategic frameworks into operational plans. With decades of specialized expertise in mobile network strategy, planning, and optimization to each of these dimensions, FL Omnitele consultancy practice is able to support you in realizing the benefits outlined in this paper.

Our services span the full breadth of the challenges described in this paper:

- AI-Era RAN Strategy
- Uplink Capacity Modeling
- 5G Revenue Planning
- Legacy Technology Migration
- Spectrum Strategy & Regulatory Engagement
- 6G Readiness Assessment
- Network Performance Audit

Whether you are navigating a legacy sunsetting program, building the business case for AI-RAN investment, or rethinking how your network planning function can drive revenue rather than merely manage cost, FL Omnitele is thrilled to support you.

Annankatu 27
00100 Helsinki
Finland

info(at)fusionlayer.com
+358 75 325 2992
www.fusionlayer.com

Sources

[1] Ericsson Mobility Report, November 2025 — Mobile data traffic growth and 5G subscription forecasts

[2] Nokia Blog: 'Physical AI: Redefining RAN and Telco Monetization' — nokia.com/blog/physical-ai-redefining-ran-and-telco-monetization/

[3] Nokia AI-RAN Overview — nokia.com/mobile-networks/ran/ai-ran/

[4] PwC Global Telecom Outlook 2025-2029 — pwc.com/gx/en/industries/tmt/telecom-outlook-perspectives.html

[5] Deloitte 2025 Telecom Industry Outlook

[6] Nokia: 'The AI Revolution: Preparing for a Surge in 5G Uplink Traffic'

[7] TelecomTV / Nokia-NVIDIA AI-RAN Summit Coverage, April 2026

[8] Dell'Oro Group: 'Level 4 Autonomous RAN — From Vision to Early Reality', 2026 — delloro.com/level-4-autonomous-ran-from-vision-to-early-reality/

[9] Dell'Oro Group: 'AI RAN — Should We Be Excited?', May 2025 — delloro.com/ai-ran-should-we-be-excited/

[10] Dell'Oro Group: 'AI-for-RAN in Focus: Key Takeaways from the Telco AI Forum', June 2025 — delloro.com

Annankatu 27
00100 Helsinki
Finland

info(at)fusionlayer.com
+358 75 325 2992
www.fusionlayer.com