



Executive Briefing

AI-RAN: NAVIGATING THE PATH FROM HYPE TO COMMERCIAL REALITY

AI-RAN is emerging as a critical area of innovation for telcos. While early pilots confirm technical feasibility, the path to commercialisation remains less clear. We explore the practical pathways for telcos to capture near-term value and prepare for long-term opportunities.



Executive summary

The artificial intelligence-powered radio access network (AI-RAN) has rapidly become one of the hottest topics in telecom. However, while early pilots are increasingly demonstrating its technical feasibility, the commercial models remain nascent. Before wider adoption can occur, telcos need to answer three critical questions that will shape the pace and scale of deployment:

- **Will telcos allow additional workloads?** Telcos remain cautious about mixing network and AI workloads on shared infrastructure, citing risks to performance, security and reliability.
- **Is compute cost still a barrier?** High-performance AI infrastructure remains expensive and power-hungry, though smaller, more efficient AI models are opening alternative paths.
- **Where is enterprise demand for AI-RAN?** Today, demand for AI workloads on RAN infrastructure is limited. Scaling adoption will require stronger enterprise engagement and clearer demonstration of value.

To assess the commercial viability of AI-RAN, we modelled three AI-RAN deployment strategies (AI-for-RAN, AI-and-RAN, AI-on-RAN), using a representative medium-sized operator with annual revenues of USD7.5 billion and a network of 15,000 cell sites. **The objective was to estimate how much financial benefit each scenario must generate over ten years to break even by year six.**

Figure 1: Commercial viability of different AI-RAN deployment strategies

AI-for-RAN	AI-and-RAN	AI-on-RAN
<p><i>Telco-centric: Applying AI internally to optimise RAN operations (e.g., spectrum management, energy efficiency, fault detection)</i></p>	<p><i>Hybrid ecosystem: Running select AI workloads on RAN infrastructure, enabling efficiency gains and creating new partnership opportunities</i></p>	<p><i>Open marketplace: Turning RAN infrastructure into a programmable, multitenant environment for AI applications</i></p>
<ul style="list-style-type: none"> • Requires peak funding of around USD150m by year three and can break even within six years – primarily through internal cost savings. • Incremental revenue uplift is modest (~1.7%), but the path is credible and low risk. 	<ul style="list-style-type: none"> • Needs roughly USD300m of peak funding by year three. • Financial returns depend not only on operational efficiency but also on monetising enterprise workloads. • By year ten, this could deliver ~4.8% revenue uplift, but success hinges on demand materialising and strong partner alignment. 	<ul style="list-style-type: none"> • The most ambitious path, requiring almost USD600m peak funding by year three. • Returns rely heavily on scaling adoption across enterprises and developers, with potential >11% revenue uplift by year ten. • However, this is highly speculative, with a significant execution risk given current uncertainty around demand.

Source: STL Partners

The modelling reinforces that AI-for-RAN is the most practical starting point, delivering tangible value today and providing a foundation for more ambitious plays. The AI-and-RAN approach offers growth potential but, in turn, requires patience, ecosystem maturity and trusted anchor customers. The AI-on-

RAN vision, while attractive in theory, remains a distant prospect, given today's risk appetite and limited evidence of enterprise-scale demand.

Recommendations

To position themselves effectively in AI-RAN, telcos should:

1. **Engage in ecosystem design:**

Telcos need to play an active role in shaping standards, governance and interoperability frameworks especially to avoid the fragmentation and limited monetisation that stunted multi-access edge computing (MEC). With hyperscalers and vendors moving fast, telcos can't afford to be passive participants.

2. **Get closer to enterprise customers:**

Before making major investments in AI-RAN, telcos need a clearer understanding of the services enterprises will need and which of these require attributes that AI-RAN can deliver, such as ultra-low latency, data sovereignty and integration. This has traditionally been a challenge for telcos, but deeper enterprise engagement is essential to ensure AI-RAN propositions align with actual demand. An early opportunity may lie in smart city deployments with local governments, where requirements for localised, low-latency services and adherence to public data governance provide a strong fit for AI-RAN capabilities. These initiatives could offer both near-term value and critical insights to inform broader enterprise strategies.

3. **Start with telco-centric AI while laying the groundwork for hybrid models:**

The initial focus should be on telco-centric AI applications that drive internal efficiency, cost reduction and operational maturity. These use cases provide a low-risk environment to test governance models, orchestration platforms and partnerships. Establishing a strong internal foundation will position telcos to expand into more ambitious hybrid models and external service offerings over time.

4. **Monitor early leaders:**

Telcos should closely track the strategies of front-runners, such as SoftBank and T-Mobile, which are setting the pace in AI-RAN deployment, orchestration and partnership models. Understanding how these leaders approach ecosystem collaboration, commercialisation and technology rollout can help telcos refine their own strategies and avoid pitfalls.

5. **Align internal teams and ownership early:**

AI-RAN success depends not only on technology but also on organisational readiness. Cross-functional alignment between RAN teams, AI specialists, enterprise sales and strategy units is critical. Telcos must break down internal silos that risk confining AI-RAN initiatives to pilot projects with no clear path to scale. Establishing a dedicated, cross-functional team with clear ownership, strong commercial focus and aligned incentives will be key to moving AI-RAN from proof of concept to value-creating.

Table of contents

Executive summary	2
Recommendations.....	3
AI-RAN is gaining momentum, but the commercial path needs to be explored further.....	6
AI-RAN is a hot topic – and early signs are promising.....	6
Efficiency benefits and, potentially, new revenue streams.....	6
Ecosystem dynamics are still forming – offering players the chance to shape outcomes.....	7
The aim of this report	7
Mapping the AI-RAN landscape: Three models, rising complexity	8
Overcoming commercial barriers: Evolving mindsets and models	11
What scale of demand is needed for a six-year payback?	13
The critical role of ecosystem partnerships.....	16
Shaping the telco role in AI-RAN	18
Our recommendations:	18

Table of figures

Figure 1: Commercial viability of different AI-RAN deployment strategies.....	2
Figure 2: Three emerging AI-RAN deployment models.....	10
Figure 3: Incremental financial impact from AI-RAN to yield cash breakeven in year six	13
Figure 4: Cumulative cash flow for different AI-RAN strategies.....	14

AI-RAN is gaining momentum, but the commercial path needs to be explored further

AI-RAN is a hot topic – and early signs are promising

AI-powered radio access networks (AI-RAN) are attracting growing interest from telcos, with early pilots demonstrating technical feasibility for both telco-centric network use cases and third-party AI workloads at the network edge. For instance, SoftBank has partnered with Red Hat to co-develop AI-RAN solutions that leverage open-source platforms, aiming to make RANs more intelligent and adaptive for next-generation services, such as 5G/6G use cases and AI-driven applications.¹ In another initiative, Red Hat and SoftBank have implemented AI-RAN in live networks to improve energy efficiency and traffic management, demonstrating how AI can directly enhance network performance and sustainability.²

Efficiency benefits and, potentially, new revenue streams

While early demonstrations of AI in the RAN are encouraging and show clear technical feasibility, the commercial path is still evolving. Initial deployments have delivered tangible benefits. In network optimisation, for instance, telcos have reported 10% gains in spectral efficiency, up to 20% higher downlink throughput³ and 15-20% reductions in energy use⁴ by applying machine learning to RAN functions. As the RAN is responsible for 75% of telcos total power consumption, adopting better resource optimisation and power use could significantly reduce their energy consumption and advance their sustainability goals.⁵ However, the broader commercial opportunity, particularly for models that extend beyond internal efficiency gains, remains in the exploratory phase. There is growing industry interest in how AI-RAN might contribute to long-term value creation via revenue growth derived from new AI services, but for now, its most proven impact lies in helping telcos run existing networks more efficiently.

“There is growing industry interest in using AI to enhance RAN operations, with early use cases including beamforming optimisation, traffic prediction and anomaly detection. These applications are becoming increasingly viable due to advances in compute platforms.”

*Researcher and principal engineer,
North American Tier 1 telco* ”

¹ <https://www.redhat.com/en/about/press-releases/red-hat-and-softbank-collaborate-develop-ai-ran-solutions>

² <https://www.redhat.com/en/about/press-releases/red-hat-and-softbank-corp-implement-ai-ran-optimize-network-performance-and-sustainability>

³ <https://www.delloro.com/ai-ran-should-we-be-excited>

⁴ https://corp.mobile.rakuten.co.jp/english/news/press/2025/0528_01

⁵ <https://www.redhat.com/en/about/press-releases/red-hat-and-softbank-collaborate-develop-ai-ran-solutions>

Ecosystem dynamics are still forming – offering players the chance to shape outcomes

At the same time, industry dynamics are shaping how value may ultimately be distributed. Ecosystem players, ranging from chip providers to AI software platforms, are playing a pivotal role by offering tools, APIs and infrastructure to support telcos on this journey. While this support accelerates progress, it also raises important questions around future ecosystem control and how value will be shared, particularly as telcos weigh whether to build their own platforms and retain greater control or rely on external providers.

Cost is another factor shaping commercial decisions. High-performance AI infrastructure can be capital-intensive and energy-demanding. At the same time, the rise of smaller, more efficient AI models, such as small language models (SLMs) that can run on central processing units (CPUs), is beginning to open alternative paths that are less expensive and less resource-intensive. As AI models continue to evolve, including the emergence of more agentic approaches that do not rely on large language models (LLMs), telcos are carefully weighing where and how to deploy AI in a way that aligns with both their cost and performance goals.

The aim of this report

There is strong agreement in the industry that AI will play a key role in the evolution of the RAN. The focus now is on identifying the most viable pathways to commercialise and operationalise it – both within existing operations and, over time, through new services and partnerships. This report aims to provide a structured exploration of AI-RAN from a commercial and operational perspective, with the following objectives:

1. To outline a range of AI-RAN implementation strategies.
2. To evaluate the associated business models for each strategy.
3. To assess the conditions under which each strategy delivers value to telcos – considering the scale and timing of investments, potential cost efficiencies and opportunities for incremental revenue.
4. To stimulate informed discussion and engagement around AI-RAN adoption and its commercial implications.

The report does not advocate for a specific AI-RAN strategy. Rather, it presents a set of preliminary analyses, assumptions and business cases intended to support telcos and vendors in evaluating AI-RAN options and shaping their strategic direction.

Mapping the AI-RAN landscape: Three models, rising complexity

There are three emerging AI-RAN deployment models, each representing a higher level of ambition, complexity and potential upside for telcos:

1. Telco-centric (or 'AI-for-RAN')
2. Hybrid ecosystem ('AI-and-RAN')
3. Open marketplace ('AI-on-RAN').

While all models are technically viable, they differ significantly in terms of control and operational requirements. These models are not mutually exclusive, meaning that telcos may evolve from one to the next over time, depending on strategy and market maturity.

1. Telco-centric model ('AI-for-RAN')

At the foundation of the AI-RAN landscape is the telco-centric model, commonly referred to as 'AI-for-RAN'. This approach is already being pursued by many Tier 1 telcos around the world. These companies have deployed AI capabilities across their live networks to drive predictive maintenance, automate resource allocation and improve energy efficiency, achieving reported savings of around 20% in some cases.

"AI for RAN aligns with the short-term telco priority of reducing costs. It improves performance using existing infrastructure, offering a low-risk internal testbed to build AI maturity without external exposure."

Senior manager – strategic technology partnerships, N. American vendor ”

The telco-centric model focuses on applying AI internally to enhance network performance, as well as operational and resource efficiency. Typical applications include dynamic spectrum management, traffic prediction, fault detection and beamforming optimisation. AI can also support dynamic resource allocation to match bandwidth and power with changing demand, improve mobility management for seamless handovers and boost energy efficiency by turning off carriers when not needed. Together, these capabilities ultimately enhance the quality of experience (QoE) for end users.

Enabling the 'AI-for-RAN' model requires a mix of centralised and distributed compute resources. Control-plane functions, such as those in the RAN intelligent controller (RIC) or the centralised unit (CU), can typically run on CPUs, while more demanding user-plane tasks at the distributed unit (DU), such as real-time scheduling or massive MIMO, may require acceleration with dedicated AI processors.

Importantly, this model is contained within the operator's own infrastructure and governance frameworks, making it operationally simpler and lower risk. It does not require exposing the network to third parties or managing multitenant environments. While this model may not create new revenue streams, it can enhance existing ones by offering more granular levels of service quality, as well as

deliver significant cost reductions and performance gains – outcomes that are increasingly critical, as telcos seek to manage rising data demands without corresponding increases in average revenue per user (ARPU).

For most telcos, this is the most practical starting point. It delivers value quickly, builds AI capabilities into existing RAN architectures and lays the groundwork for more ambitious future models.

2. Hybrid ecosystem model ('AI-and-RAN')

Building on the telco-centric foundation, the hybrid model introduces a shared infrastructure approach. Here, telcos continue to operate their core and RAN network functions but also begin hosting additional AI workloads, which may include their own specialised models such as SLMs, or solutions provided by third parties. For instance, telcos might run a partner's computer vision model for industrial robotics on the same edge infrastructure used for local RAN functions.

The hybrid strategy would be typically structured around formal commercial agreements, often with large enterprises, governments or vertical solution providers, allowing telcos to monetise spare compute capacity and provide value-added services to enterprise customers. Technically, this requires more advanced orchestration capabilities. Telcos must be able to isolate workloads, guarantee performance, and maintain network reliability and security, even as non-network applications share infrastructure. It also introduces new commercial and operational challenges: partner onboarding, service-level agreement (SLA) enforcement, data governance and support functions must all mature significantly.

However, the payoff can be compelling. The hybrid model opens the door to incremental revenues via partner access fees, joint solutions or data-as-a-service (DaaS) propositions. It also enables telcos to co-create sector-specific applications, such as AI-assisted public safety, financial fraud detection and autonomous vehicle coordination, on infrastructure they already operate. This makes the model particularly relevant in high-value, high-trust environments where latency, locality and data sovereignty matter.

For telcos seeking to diversify revenue, the hybrid model offers a commercially attractive next step.

3. Open marketplace model ('AI-on-RAN')

The most ambitious vision is the 'AI-on-RAN' model. While there are significant technical prerequisites, STL Partners' view is that commercialisation will ultimately depend on building a true marketplace approach on top of the technical foundations. As such, this model would position the telco not just as a network operator but as a marketplace provider, enabling third-party developers to deploy and monetise AI applications directly on RAN infrastructure.

Under this approach, the telco's RAN infrastructure becomes a programmable, multitenant environment. Developers can onboard themselves via self-service tools, access standardised APIs and offer services to a long tail of enterprises and consumer customers. Use cases could span immersive entertainment, smart city digital twins, remote diagnostics and on-demand robotics control.

Technically, the marketplace model demands a cloud-native architecture with strong workload isolation, dynamic resource allocation and zero trust network. It also requires capabilities beyond the network domain: marketplace orchestration, billing and settlement systems, developer support and governance policies are all essential to sustaining a viable marketplace.

This model has the highest potential upside in terms of revenue diversification and ecosystem growth. If successful, it could enable telcos to participate in entirely new categories of value creation, powered by developer innovation and supported by network-native capabilities, such as near real-time contextual awareness and low-latency delivery.

“AI and RAN has the biggest potential, especially with all the new players and platforms in the market. Operators must join this ecosystem of players to stay competitive and scale globally.”

Principal fellow, Asian Tier 1 telco ”

However, it also comes with the greatest uncertainty. Success depends on the emergence of a vibrant third-party developer ecosystem, alignment with enterprise IT environments and a clear business case for deploying workloads directly on RAN infrastructure rather than in central clouds or on devices.

Figure 2: Three emerging AI-RAN deployment models

Aspect	Strategy 1: Telco-centric	Strategy 2: Hybrid ecosystem	Strategy 3: Open marketplace
Control model	Exclusive telco control	Telco primary with partner access	Multitenant platform model
Third-party access	Limited API access	Formal partner agreements	Open on-demand access
Value proposition	Network optimisation	Balanced network and partner solutions	Edge compute marketplace
Revenue model	Enhanced telco services	Service fees + revenue sharing	Platform fees + transaction cut
Deployment complexity	Moderate	High	Very high
Data governance	Centralised with telco	Federated with partners	Distributed multilevel
Regulatory challenges	Minimal	Moderate	Significant
Competitive position	Traditional telco+	Ecosystem player	Platform provider
Investment and ROI potential	Moderate	High	Very high

Source: STL Partners

Each model could build upon the capabilities of the last, requiring progressively more advanced technology, operational maturity and organisational alignment. Rather than choosing one model definitively, many telcos may adopt a phased approach – starting with internal optimisation, experimenting with hybrid deployments and assessing market conditions before committing to a full marketplace strategy. Timing, partner maturity and alignment with broader cloud and AI trends will influence how far and how fast telcos move along this curve. Having said that, the open marketplace does require a fundamentally different commercial approach to the hybrid ecosystem. Where the latter is about managing a small number of very large customers tied to bespoke commercial agreements and technical implementations, the open marketplace is about managing a long tail of customers via standardised and automated commercial and technical processes and agreements. Success with the hybrid ecosystem may preclude moving to the open marketplace.

Overcoming commercial barriers: Evolving mindsets and models

While the technical feasibility of AI-RAN is becoming increasingly clear, telcos still face a range of commercial and organisational hurdles that must be addressed before wider adoption can occur. In particular, three critical questions are emerging that will shape the pace and scale of deployment.

1. Will telcos allow additional workloads?

Many telcos are still cautious about opening their network infrastructure to additional workloads. This caution is shaped in part by lessons from MEC. While the early promise of MEC was compelling, in practice, uptake was limited – partly because enterprise demand didn't materialise at scale, and in part because telcos lacked the tools and governance models to support new applications at the network edge with the necessary confidence.

That experience continues to influence thinking around AI-RAN, particularly for hybrid and open models. Running AI workloads on the same infrastructure as live RAN functions introduces perceived risk – from performance unpredictability and data leakage to compliance concerns. For operational teams used to strict SLAs and deterministic performance, introducing less predictable workloads can be uncomfortable.

Today, we see a split in operator strategies: some insist on isolation – keeping any AI workloads on separate, standalone infrastructure – to minimise risks. Others are exploring integrated approaches (shared infrastructure) to improve utilisation and economics, but only slowly and with trusted partners. The lack of common frameworks and standards exacerbates this caution. Unlike hyperscalers which confidently run multitenant clouds with strict isolation, telcos do not yet have a mature platform to manage mixed criticality workloads across their RAN infrastructure. They would need strong safeguards – security, quality assurance (QA), fail-safe mechanisms – and a clear financial upside to overcome internal resistance to opening the network.

“Operators are unlikely to open infrastructure to third parties until they are confident in robust orchestration and clear boundaries between RAN and AI workloads.”

*Vice president, RAN operations,
North American vendor*

”

Ultimately, for telcos to allow broader deployment of AI workloads on RAN infrastructure, they will need to be confident in their ability to enforce guardrails around security, performance and policy, and see a compelling financial case for doing so. Until that foundation is in place, most deployments are likely to remain selective and use case-specific.

2. Is compute cost still a barrier?

AI workloads require more compute than traditional RAN functions – particularly when dealing with unstructured data or real-time applications deployed alongside user plane functions. This introduces new cost considerations – both capital expenditure (capex) for aspects such as graphics processing units (GPUs), memory and networking and operational expenditure (opex) on power, cooling and

maintenance. While chip prices are gradually falling and more efficient edge-grade accelerators are becoming available, the economics of AI at the edge remain a question.

Nvidia's L4 GPUs, priced at around USD4,000, are intended to provide a cost-effective entry point for edge inference tasks, enabling a 'start small, scale smart' strategy that allows telcos to experiment without large upfront investments. In contrast, high-performance workloads may still require more powerful chips, such as Nvidia's Grace Hopper and Blackwell, which can cost upwards of USD125,000. As such, a hybrid infrastructure model is emerging as a pragmatic solution where CPUs handle orchestration and standard RAN functions, while GPUs are used for AI-specific workloads. This way, telcos can ensure that high-cost compute is only used where performance gains or cost savings justify it.

"Operators can start small with L4s to run lighter AI tasks and scale over time, avoiding upfront investment in high-end chips such as Grace Blackwell. It's worth noting, though, that a single Grace Blackwell can support up to 20 sites, compared to one site per CPU, which can significantly improve cost efficiency at scale."

Principal fellow, Asian Tier 1 telco ”

Another consideration is supplier dependency. Today's AI-RAN ecosystem is heavily influenced by a small number of chip vendors. While these suppliers play a key role in accelerating adoption, some telcos are wary of repeating past mistakes where proprietary hardware and software limited flexibility and drove up long-term costs. This is prompting interest in open alternatives and multi-vendor strategies, but maturity varies.

3. Where is enterprise demand for AI-RAN?

Finally, the question of demand remains central, particularly for hybrid and open models that depend on enterprise and developer uptake. While telcos widely agree that enterprise interest in edge AI is growing, it is less clear whether there is strong, scalable demand for AI workloads to run specifically at the RAN edge.

So far, demand is concentrated in a handful of verticals, typically in scenarios where low latency, data locality or privacy is critical. Examples include AI-powered robotics in industrial environments, secure video analytics for government use and real-time logistics tracking in retail. This doesn't mean demand won't grow. As enterprises look to embed AI into operational workflows and as use cases become more latency- or data-intensive, the value of edge AI will increase.

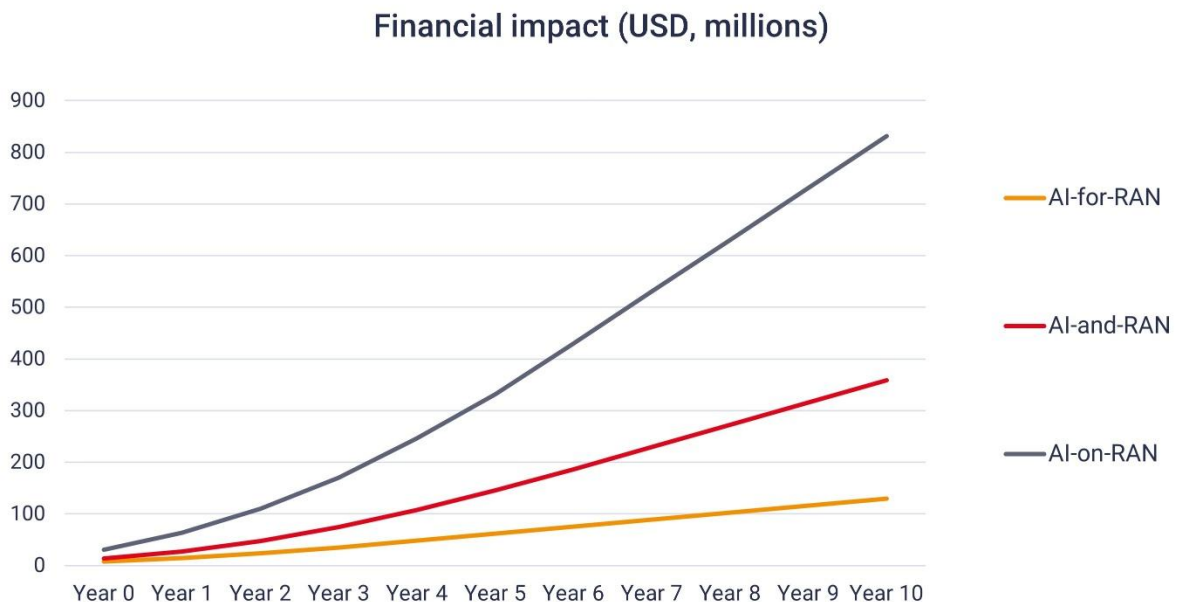
Telcos should treat this as a reason to proceed cautiously. Internal use cases, such as optimising RAN performance and automating operations, offer a clearer and more immediate return – and should be the priority. At the same time, working closely with a small number of enterprise partners can help telcos validate which use cases truly benefit from AI at the edge and test the infrastructure, orchestration and commercial models required to support them. In essence, telcos need to be assured that there are enough use cases where AI compute found at a regional data is not sufficient and customers are willing to pay a premium for AI deep in the network.

What scale of demand is needed for a six-year payback?

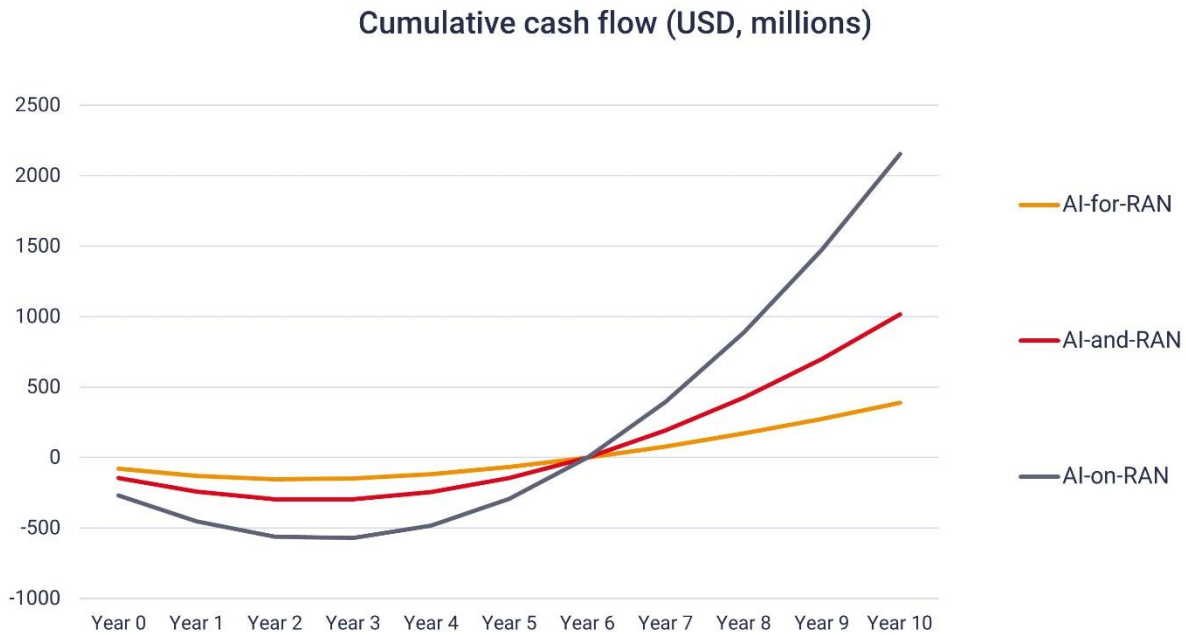
To assess the commercial viability of different AI-RAN deployment strategies, we modelled the three scenarios, using a representative medium-sized operator with annual revenues of USD7.5 billion and a network of 15,000 cell sites. The objective was to estimate how much financial benefit each scenario must generate over ten years to break even by year six. In other words, this is not a forecast for each model but an evaluation of 'what needs to be true' for each of these approaches to achieve a six-year payback.

The analysis is based on a full cash flow model, incorporating capital investments, operational costs and anticipated efficiency-driven cost savings. Capital investment includes edge AI hardware, orchestration software, integration, security and ecosystem enablement where applicable. These investments are weighted towards the early years, particularly in scenarios involving third-party partnerships. Operational costs cover AI operations, model training, data centre usage, compliance, partner management and commercial overheads, with cost assumptions scaling in line with revenue. The model also accounts for AI-driven savings in areas such as automation, predictive maintenance and fault management, which are expected to grow over time.

Figure 3: Incremental financial impact from AI-RAN to yield cash breakeven in year six



Source: STL Partners

Figure 4: Cumulative cash flow for different AI-RAN strategies

Source: STL Partners

- The **telco-centric AI-RAN strategy** emerges as the most straightforward route to break even. With peak funding of around USD150 million in year three and financial returns derived mainly from internal cost savings, it reflects a low-risk, operationally focused strategy. This outcome is not unexpected, given that 'AI-for-RAN' is already delivering tangible value, with telcos using it to reduce costs through automation, fault prediction and network optimisation. Early deployments demonstrate that these benefits are real and achievable. Incremental revenue in this model is modest – representing a 1.7% uplift on an operator's base revenue.
- The **hybrid AI-RAN ecosystem strategy** requires peak funding of approximately USD300 million in year three. This model builds on proven operational benefits while adding selective monetisation opportunities with enterprise customers. Incremental revenue is USD360 million in year ten, which represents an increase of 4.8% on an operator's base revenue. Although more commercially ambitious, it suggests that there could be a path to growth for telcos seeking to expand beyond pure internal efficiencies. This, of course, is dependent on demand for AI-RAN workloads.
- The **open AI marketplace strategy** assumes a broader role for the operator in enabling external developer and enterprise ecosystems. This requires a substantial uplift in peak funding requirements – hitting nearly USD600 million in year three. Success is much more dependent on generating incremental revenue – around USD850 million in year ten, representing an 11% uplift on base revenue. Its success relies on scaling adoption across multiple customer segments and achieving a leading position in the AI services market. While AI inferencing requirements are

increasing edge compute demand, it is not clear at all that there will be this level of demand for AI-RAN.

Overall, a telco-centric approach ('AI-for-RAN') provides operators with a credible and relatively low-risk starting point to capture AI-driven value. With thoughtful implementation, this model also preserves the option of pursuing more ambitious strategies in the future. At present, enterprise and developer demand for AI-RAN workloads remains limited, so it makes sense for telcos to focus on assessing the technical and commercial viability of running telco and enterprise workloads on shared infrastructure. Alongside this, they should continue exploring opportunities to secure large enterprise and public sector customers for AI-RAN solutions.

Currently, however, it looks too early to consider a hybrid ecosystem launch – although things could change fast here, given how quickly AI initiatives are developing. The open marketplace strategy also looks attractive in theory, but current lack of known enterprise and developer demand, coupled with most telcos' limited investment budget and risk appetite, means that this remains a distant dream.

The critical role of ecosystem partnerships

To succeed with AI-RAN, telcos will need to engage a broad ecosystem of partners. AI-RAN is not merely a network upgrade; it represents a significant shift towards a more software-defined, AI-enabled network architecture. Delivering this at scale will require capabilities that most telcos do not have in-house, particularly in areas such as AI lifecycle management, orchestration and multitenant workload management. Success will depend on building partnerships that help telcos address these capability gaps while enabling them to remain competitive in a rapidly evolving market.

Key partners will include (though not be limited to):

- **Network vendors** – which will supply AI-integrated RAN equipment and develop the software frameworks necessary for enabling AI workloads at the network edge. These vendors play a pivotal role in embedding AI functions such as automation, network optimisation and intelligent control within existing infrastructure.
- **Hyperscalers** – which can provide critical expertise in cloud infrastructure, orchestration and workload management. They also offer platforms that support developer engagement and multitenant AI operations. These capabilities are essential for enabling flexible AI deployment models and supporting third-party developer ecosystems. However, telcos also have the option of building and operating their own cloud, particularly in contexts where sovereignty and data control are key considerations.
- **Chipmakers** – which design and manufacture the specialised processors required for AI workloads, including GPUs, neural processing units (NPUs) and AI accelerators. These components are key to supporting both AI model training in central data centres and low-latency inference at the edge of the network. The performance, efficiency and availability of this hardware will have a direct impact on the viability of AI-RAN solutions.
- **Independent software vendors (ISVs)** – which bring valuable AI applications and services to the market. These may range from network-centric tools that optimise performance to industry-specific solutions for sectors such as manufacturing, retail, healthcare and government. ISVs offer telcos a way to expand their service portfolio, especially when pursuing hybrid or marketplace business models aimed at enterprise customers.
- **Systems integrators (SIs)** – which play a critical role in bringing together hardware, software and cloud capabilities into cohesive, deployable solutions. They can help telcos manage complex multi-vendor environments, accelerate deployment timelines and ensure interoperability across the AI-RAN ecosystem.

This ecosystem will likely continue to expand as AI-RAN technologies mature. New players, including AI startups, specialist integrators and open-source communities, are expected to emerge with innovative solutions and approaches.

However, while partnerships are essential to accelerate AI-RAN deployment and enable new commercial models, they also present strategic considerations. An important consideration is the degree of control telcos will retain over critical functions within the AI-RAN value chain. For instance, if areas such as orchestration, AI management and developer engagement are primarily led by external partners, telcos may find it harder to capture the full value of their AI investments. As such, telcos must engage proactively with ecosystem partners and seek to establish relationships that are not only collaborative but also aligned with their long-term strategic goals. This includes negotiating clear terms around data ownership, platform governance and monetisation rights, ensuring that telcos retain influence over key operational and customer-facing domains.

“Partnerships will be central to delivering AI-RAN models. However, control of the customer relationships will dictate value capture in partnerships, so telcos need to make sure they understand enterprise customer needs.”

Senior manager – strategic technology partnerships, N. American vendor

As such, telcos must engage proactively with ecosystem partners and seek to establish relationships that are not only collaborative but also aligned with their long-term strategic goals. This includes negotiating clear terms around data ownership, platform governance and monetisation rights, ensuring that telcos retain influence over key operational and customer-facing domains.

It will also be important for telcos to invest in building internal competencies, particularly in orchestration, AI service integration and partner ecosystem management. Developing these capabilities will help reduce over-reliance on any single partner and allow telcos to remain agile as the market evolves.

Ultimately, while ecosystem partnerships are a critical enabler for AI-RAN, telcos must balance the benefits of collaboration with the need to safeguard their strategic position. Those that succeed will be those who engage early, shape partnerships deliberately and invest in the capabilities required to remain active participants in the AI-RAN value chain.

Shaping the telco role in AI-RAN

As the industry moves towards AI-native network architectures, telcos have a critical window of opportunity to shape their role in this evolving landscape. This is not solely about short-term benefits but also about securing their long-term relevance in a market that is undergoing structural change. The shift towards AI-driven networks is inevitable – and telcos that delay engagement risk losing influence over where and how future value is created.

Our recommendations

To position themselves effectively in AI-RAN, telcos should:

- 1. Engage in ecosystem design:**

Telcos need to play an active role in shaping standards, governance and interoperability frameworks especially to avoid the fragmentation and limited monetisation that stunted MEC. With hyperscalers and vendors moving fast, telcos can't afford to be passive participants

- 2. Get closer to enterprise customers:**

Before making major investments in AI-RAN, telcos need a clearer understanding of the services enterprises will need and which of these require attributes that AI-RAN can deliver, such as ultra-low latency, data sovereignty and integration. This has traditionally been a challenge for telcos, but deeper enterprise engagement is essential to ensure AI-RAN propositions align with actual demand. An early opportunity may lie in smart city deployments with local governments, where requirements for localised, low-latency services and adherence to public data governance provide a strong fit for AI-RAN capabilities. These initiatives could offer both near-term value and critical insights to inform broader enterprise strategies.

- 3. Start with telco-centric AI while laying the groundwork for hybrid models:**

The initial focus should be on telco-centric AI applications that drive internal efficiency, cost reduction and operational maturity. These use cases provide a low-risk environment to test governance models, orchestration platforms and partnerships. Establishing a strong internal foundation will position telcos to expand into more ambitious hybrid models and external service offerings over time.

- 4. Monitor early leaders:**

Telcos should closely track the strategies of front-runners, such as SoftBank and T-Mobile, which are setting the pace in AI-RAN deployment, orchestration and partnership models. Understanding how these leaders approach ecosystem collaboration, commercialisation and technology rollout can help telcos refine their own strategies and avoid pitfalls.

- 5. Align internal teams and ownership early:**

AI-RAN success depends not only on technology but also on organisational readiness. Cross-

functional alignment between RAN teams, AI specialists, enterprise sales and strategy units is critical. Telcos must break down internal silos that risk confining AI-RAN initiatives to pilot projects with no clear path to scale. Establishing a dedicated, cross-functional team with clear ownership, strong commercial focus and aligned incentives will be key to moving AI-RAN from proof of concept to value-creating.

PARTNERS



Research



Consulting



Events