GSMA Intelligence

Telco AI:
State of the Market, Q1 2025

AI and networks

March 2025

gsmaintelligence.com

@GSMAi

GSMA™

The GSMA is a global organisation unifying the mobile ecosystem to discover, develop and deliver innovation foundational to positive business environments and societal change. Our vision is to unlock the full power of connectivity so that people, industry, and society thrive. Representing mobile operators and organisations across the mobile ecosystem and adjacent industries, the GSMA delivers for its members across three broad pillars: Connectivity for Good, Industry Services and Solutions, and Outreach. This activity includes advancing policy, tackling today's biggest societal challenges, underpinning the technology and interoperability that make mobile work, and providing the world's largest platform to convene the mobile ecosystem at the MWC and M360 series of events.

We invite you to find out more at www.gsma.com

Authors

Peter Jarich
Head of GSMA Intelligence

Tim Hatt
Head of Research and Consulting

Sayali Borole
Senior Analyst

GSMA Intelligence

GSMA Intelligence is the definitive source of global mobile operator data, analysis and forecasts, and publisher of authoritative industry reports and research. Our data covers every operator group, network and MVNO in every country worldwide – from Afghanistan to Zimbabwe. It is the most accurate and complete set of industry metrics available, comprising tens of millions of individual data points, updated daily.

GSMA Intelligence is relied on by leading operators, vendors, regulators, financial institutions and third-party industry players, to support strategic decision-making and long-term investment planning. The data is used as an industry reference point and is frequently cited by the media and by the industry itself.

Our team of analysts and experts produce regular thought-leading research reports across a range of industry topics.

www.gsmaintelligence.com

info@gsmaintelligence.com

GSMA Intelligence

AI for networks versus networks for AI

Peter Jarich
Head of GSMA Intelligence

When we began our State of the Market series in 2024, the goal was to provide a view of the many ways operators were innovating with AI, and flag some dynamics that require more attention if operators are to fully exploit the opportunity that AI presents.

The series has covered a lot of ground, including a deep dive into open versus proprietary large language models (LLMs), the role of energy pressures in scaling the use of AI, and a framework for how to think holistically about the return on investment from AI. This has been accompanied by real-world examples of telco AI in action.

This latest iteration includes one of the most important industry dynamics: AI for networks versus networks for AI.

AI has been used by telecoms operators in their network operations for years – whether to guard against threats, minimise energy usage or optimise RAN performance. As AI's use proliferates, operators are in a unique position to enable its growth through network investment and evolution. Providing a foundation from which AI can thrive will require investment – network capacity expansion, workload resilience, and orchestration between cloud and edge resources. The fundamental role telecoms networks play in AI's success provides a reminder of how critical operators are in enabling innovation – and the monetisation opportunity this presents.

About the series

This is the final report in a four-part series on AI strategy in telecoms. This edition examines the impact of AI on telecoms networks, and the flipside of how network capabilities can be leveraged for AI.

Telco AI: State of the Market quarterly series: research approach

Market context	Leader profiles	Deep dives

• A recap of critical industry developments and implications
• Putting industry progress in context, along with AI and security synergies
	• Case studies of telco AI in action
	• Understanding best practice for the sector
		• Deep dives into key topics
		  • AI data and systems
		  • energy efficiency
		  • strategy and culture
		  • network strategy

In tandem with the report series, GSMA Intelligence has developed an AI benchmark to track AI use and implementation in the telecoms sector.

GSMA Intelligence

01 Executive summary

Summary

01

AI-driven network efficiency and cost optimisation

AI is making networks intelligent, efficient and cost effective. Use of AI has evolved from automating tasks to optimising larger systems. In network operations, predictive maintenance, self-healing capabilities and intelligent traffic management have reduced downtime and improved cost efficiencies. Taking a step further, genAI enhances automation by enabling real-time decision-making, optimising resource allocation and personalising customer interactions, making networks more adaptive.

02

AI for networks: managing traffic and bandwidth demand

Networks are the core of the operator business. With 5G adoption and AI-powered applications driving exponential traffic growth, networks must handle increasing bandwidth demand. AI solutions can help networks keep up through dynamic traffic routing, intelligent load balancing and automated congestion management to optimise performance while maintaining service quality.

03

AI for networks: energy and sustainability

The RAN is the most energy-intensive part of telecoms infrastructure, at 80–85% of total consumption – a cost line that has remained stubbornly high. AI-powered RAN Intelligent Controllers and traffic steering can enhance resource allocation, automate fault detection and support sustainability efforts by reducing power consumption in network operations. Integrating AI into RAN operations can harvest downstream energy efficiencies, offsetting the AI-driven energy rises from large language models (LLMs) and other model training in data centres.

04

Networks for AI: inferencing and edge

Smarter, more agile network infrastructure is essential to support AI applications. AI requires significant processing and computing power. Cloud integration enables AI training and large-scale computing, while edge computing reduces latency and operational costs by processing data closer to its source. The edge is also strategically important to allow for distributed compute, which will become a key competitive advantage because of the huge projected rise in data traffic from AI.

05

Networks for AI: new revenue models

AI inference opens up new use cases in areas such as industrial robotics, digital twins, IoT and VR/AR, allowing operators to create new revenue streams with enterprise customers. GPU-as-a-service, AI model licensing and localised AI infrastructure all offer monetisation opportunities for operators. Local AI processing is also seeing growing demand. Meanwhile, distributed computing reduces energy costs, making AI adoption both a revenue driver and cost-saving strategy for enterprise clients.

AI in numbers

80%

Growth imperative

Improving customer experience and supporting new services beat opex and capex savings as the top strategic network transformation priorities by a four-to-one margin, with 80% most focused on the revenue-generating activities.

53%

AI for maintenance and planning

53% of surveyed operators believe genAI will have the greatest impact on network planning and optimisation, followed by network troubleshooting and predictive maintenance. Maintenance use cases include the use of agents and natural language to keep networks and services running. Planning and optimisation are generally supported by core AI technologies, with room to expand with genAI.

3×

AI investment implications

15% of operators expect added data traffic to be the biggest network impact from genAI. Three times that percentage believe internally focused impacts will dominate (e.g. planning/optimisation, improved software development and predictive maintenance).

62%

Expertise and experience paramount

Across many AI investment requirements, 63% of operators prioritised AI investment in upskilling, followed closely by network capacity (62%). The former points to the requirement for constant training and expertise development while using AI. The latter highlights recognition that AI will drive more (and new types of) traffic across networks.

GSMA Intelligence

02 Market context

Recent developments: news flow

Operator-led

Verizon unveils AI strategy for enterprises

The Verizon AI Connect offering blends the operator's fibre infrastructure (which includes its long-haul, metro fibre, local fibre and optical networks) with its power, space and cooling capabilities.

Jan 2025

Orange and Mistral AI join forces in Europe AI push

Orange has teamed up with Mistral AI to boost adoption of AI in Europe, with the pair pledging to work together to design the best network infrastructure to accommodate the technology.

Feb 2025

Nokia Bell Labs and STC target network boost with genAI

Nokia's Bell Labs Consulting has teamed up with STC Group to unveil a genAI offering designed to reduce costs and time to market for new services launching on the operator's network.

Dec 2024

SKT unveils regional AI hub plan

SKT plans to build an AI infrastructure superhighway, starting with the launch of a cloud-based GPU-as-a-service through a partnership with Lambda.

Nov 2024

SoftBank teams up with Nvidia on supercomputer, AI-RAN

SoftBank plans to use Nvidia's Grace Blackwell platform for its next supercomputer, to run extremely compute-intensive workloads. It conducted the world's first AI-5G RAN outdoor trial.

Nov 2024

Ecosystem-led

Korea allocates $32m to nurture AI talent

South Korea's Ministry of Science and ICT called for applicants for the 2025 AI Star Fellowship Programme, part of a five-year initiative to support top-tier, early-career researchers in AI.

Jan 2025

European tech firms come together to launch sovereign edge cloud

The Virt8ra edge cloud will enable the development of next-generation use cases that require ultra-low latency.

Jan 2025

Extreme Networks readies AI enterprise platform

US-based vendor Extreme Networks unveiled a platform to simplify and automate product licensing for enterprises and channel partners by using an AI-based software suite.

Dec 2024

Verizon and Nvidia power AI across 5G private networks, MEC

Nvidia is working with Verizon to develop a range of AI applications that run on the operator's 5G private network in tandem with its private mobile edge compute (MEC) service that puts applications closer to users.

Dec 2024

Ericsson and Grameenphone target AI advances

Grameenphone and Ericsson allied on AI and automation research, seeking to foster growth and advance digital transformation. With AI and automation, operators can better respond to the evolving needs and increasing demand on networks.

Nov 2024

Source: GSMA Intelligence and Mobile World Live, based on company announcements

AI in the news: implications

Significant developments around AI-driven networks	• There have been a number of announcements of operators using AI for networks to optimise and automate network operations. Operators are investing billions of dollars in their networks. Looking for ways to leverage new technology innovations to shave costs is an obvious next step. • Developments indicate strong momentum behind using AI to optimise network operations and building networks designed to support AI workloads.
Networks for AI	• If networks are to be an AI asset, operators will need to think more about their capabilities versus AI demands. • The growing number of partnerships indicates progress towards AI-driven infrastructure, with networks becoming increasingly capable of supporting and benefiting from AI workloads. For instance, Nokia, Ericsson and Extreme Networks are pushing the boundaries by developing both hardware and software solutions.
Edge computing and AI at the network periphery	• There is a growing focus on deploying AI and edge computing at the network edge, driven by the need to reduce latency and improve real-time decision-making, especially for time-sensitive applications such as IoT, autonomous vehicles and gaming. • With AI running closer to end users, networks will have to become more responsive and capable of handling the increasing volume of data generated by connected devices.
Rise in local initiatives, building capacity	• The rise of autonomous edge clouds in Europe and AI talent initiatives in countries such as South Korea signal a shift towards localising AI infrastructure, ensuring that national and regional ecosystems are prepared to harness its potential while addressing issues such as data privacy, cybersecurity and talent development.

03 Leader profiles: telco AI in action

AI in action: Nokia AI and analytics solutions

Nokia's AVA AI and analytics solutions, enhanced with genAI, are positioned to change how operators access and use data. By integrating genAI, the solutions, such as AVA Data Suite, empower employees with autonomous agents that facilitate self-service data access, democratising knowledge retrieval and accelerating decision making. Even without SQL skills, operators can interact with databases using natural language queries to ask the genAI assistant for insights on network quality and service.

AVA Data Suite: supporting autonomous networks

- **The challenge.** With growing numbers of users and internet/data-centric applications, operators and other service providers face a challenge in managing and optimising increasingly intricate networks. As networks evolve, manual operations become less feasible, leading to higher operational costs, increased energy consumption and a potential decline in service quality. The need for efficient, automated solutions to handle vast amounts of data and ensure optimal network performance is paramount.
- **The solution.** The AVA Data Suite is designed to help accelerate the journey to autonomous networks. The suite employs a data mesh architecture for automated, domain-agnostic data standardisation, ensuring operators have timely access to necessary data. It features a telco-centric, self-service AI studio equipped with an MLOps/LLMOps framework and pre-packaged AI models, enabling operators to develop their own AI and genAI use cases. The cloud-agnostic nature of the solution allows seamless integration with existing ecosystems and environments.

Benefits

Nokia AVA Data Suite streamlines data access for real-time insights, freeing data scientists to focus on developing AI use cases. Without it, operators can face higher costs, slower insight gathering and longer data preparation time.

- **Speeding up AI implementation in networks.** Through Nokia's testing, with AVA Data Suite's reusable data products, the data preparation process in the AI/ML lifecycle can be completed in 3–4 weeks – around 70% faster in terms of time to value than in a typical project (four months).
- **Comprehensive network and subscriber visibility.** The solution transforms raw data with domain knowledge and intelligence into advanced and cognitive data products that provide high-value actionable insights and a foundation for AI use cases. Users can build and share custom data products, incorporating external sources. Robust governance ensures data integrity, synchronisation, reliability and security through a data catalogue and lineage mechanisms.

AI in action: Nokia AI and analytics solutions (continued)

Impact for operators

Reported impacts from the Nokia AVA AI and analytics solutions for operators include enhanced productivity, improved customer satisfaction and new revenue growth. Areas where the solutions are having an effect include the following:

- **Using multi-source data for actionable insights.** The solution integrates data from multi-vendor networks (both fixed and mobile) alongside subscriber data, including devices and apps. It transforms raw data into cognitive data products that deliver meaningful and actionable insights. The insights serve as a foundation for AI-driven use cases such as automating operational workflows, enhancing quality of experience, reducing energy consumption and optimising network performance.
- **Enhanced real-time customer experience insights.** Using AI-driven AVA Experience Prescriptions, AVA Data Suite enables real-time network optimisation. The solution uses unsupervised anomaly detection and root-cause analysis to provide actionable recommendations that can improve voice, video and gaming experiences. For instance, one operator achieved a 60% reduction in Netflix buffering and a 15% decrease in YouTube playback delays.
- **Empowering employees with autonomous agents.** The solution empowers employees with autonomous agents that facilitate self-service data access. Employees can now use natural language queries instead of SQL, speeding up decision-making processes. Even employees without SQL knowledge can interact with databases intuitively. For example, they can ask the assistant, “Provide a dashboard of quality of experience in region X” or “Is there an issue with subscriber X’s Wi-Fi?” to obtain insights quickly and efficiently.

150+

More than 150 operators globally are using the AVA solutions. The solutions have been instrumental in reducing energy consumption by optimising power usage in both active and passive network equipment and enhancing network performance.

80%

For China Mobile, use of the solution resulted in an 80% reduction in knowledge acquisition time, a 72% increase in efficiency of data analysis and €6.5 million in annual savings for network operations. (Source: Nokia)

AI in action: LG Uplus AI Agent initiatives

In July 2024, LG Uplus shared its 'All in AI' enterprise services strategy, with the aim to drive $1.4 billion in annual revenues by 2028. The operator is set to invest in becoming an AI-focused company. It plans to leverage AI's capabilities to enhance services and deliver greater value to its customers, envisioning a future with autonomous agents. This has led to the development of AI agent solutions including the ixi-O AI Call Agent, designed to address the evolving communication needs of Gen Z, and the ixi-GEN based Work Agent, which aims to optimise and streamline complex network operations.

ixi-O Call Agent (networks for AI)

The challenge. The Gen Z demographic group generally makes fewer traditional phone calls. Data from LG Uplus showed that longer calls were usually only made with select people, indicating phone calls for the Gen Z group are about intimacy. LG Uplus claims that existing services have failed to recognise this behaviour.

The solution. Implemented in 2024, ixi-O is designed to offer a secure and seamless calling experience specifically for the Gen Z demographic group. Based on user feedback and research, it aims to re-imagine phone calls through three core features:

- **AI call reception,** which allows AI to answer calls, record conversations and provide text transcripts.
- **Real-time text communication,** with live transcriptions enhancing comprehension and engagement.
- **Voice phishing detection,** which alerts users to potential scams in real-time, ensuring improved trust and security for phone calls.

Recognising Gen Z's comfort with text-based communication, the operator plans to evolve the agent into a personal agent, capable of making smart recommendations and planning schedules, for example. It will take care of a range of daily tasks through the use of genAI.

The key solutions and technology behind ixi-O include on-device AI, which enhances privacy by processing data locally while reducing network load and server costs. The agent also uses ixi-GEN and LLMs from big tech where it needs greater compute capabilities for specific tasks. Additionally, the Hybrid Communication Platform seamlessly integrates with existing phone numbers by combining mVoIP and VoLTE, providing a consistent and reliable calling experience.

AI in action: LG Uplus AI Agent initiatives (continued)

ixi Work Agent (AI for networks)

The challenge. LG Uplus struggled to manage its increasingly complex network operations. Dealing with multiple systems, vast information retention and adherence to standardised procedures can lead to operational inefficiencies and staff frustration .

The solution. ixi-GEN based Work Agent, implemented in 2024, was developed to streamline and optimise network operations. The aim is to improve efficiency, enhance network stability and boost customer experience through the following:`

- An automated decision-making and action agent detects network issues and takes corrective action autonomously.
- A decision support co-pilot provides detailed analysis and visualisation in decision-making.
- A network operations knowledge-base chatbot offers instant access to operational knowledge, improving efficiency.

Deep learning algorithms detect early signs of equipment failure, enabling proactive maintenance and preventing potential network disruptions. Going forward, the agent will eventually oversee the entire process autonomously – from issue detection to taking actions.

Implementation approach. ixi Work Agent integrates ixi-GEN with LangChain's agent framework, combining internal knowledge sources such as standard operating procedures and manuals with API-driven data retrieval and task execution. To ensure safety and risk mitigation, a verification process is in place to prevent incorrect actions, with critical operations requiring operator confirmation. Additionally, closed network development enhances security by leveraging in-house ixi-GEN.

100,000

ixi-O saw more than 100,000 downloads within 10 days of launch, and its popularity continues to grow.

4×

Since the introduction of ixi-O, the operator has seen a fourfold increase in iPhone sales.

ixi-GEN has significantly improved operational efficiency, reduced task lead times and streamlined issue resolution. It has also contributed to a more stable network.

ixi-O as a personal AI assistant will shift more functions to the device (from cloud) for enhanced personalisation and security.

The operator plans to expand AI agents to cover a broader range of network operation tasks. Interoperability between agents will be enhanced for better analysis and collaborative decision-making.

GSMA Intelligence

04 Deep dive: AI for networks

AI = efficiency[2]?

- **Automation for efficiency.** Use of AI for networks is not new. Operators have evolved from simply automating processes to optimising operations and exploring new revenue streams. Core AI technologies are transforming traffic management, predictive maintenance and network self-healing, leading to reduced downtime. This is driving significant operational cost savings. Meanwhile, genAI is augmenting automation through real-time decision-making, and personalised customer interactions, making networks more adaptive and efficient.
- **GenAI for cost optimisation.** As operators invest billions in network infrastructure, leveraging technological innovations to reduce costs is essential. GenAI plays a crucial role in augmenting existing capabilities and introducing new functionality, including traffic management, billing, operations and maintenance. These advancements drive significant cost savings and operational efficiencies, helping operators maximise their investment.
- **Towards a customer-centric ecosystem.** While operators initially focused AI investments on improving operational efficiencies, future strategies will increasingly balance this with efforts to enhance customer experience and generate revenue. AI's ability to analyse customer behaviour and optimise services will make networks more efficient, more profitable and ultimately more customer-centric. As traffic demand continues to rise, this shift in focus will enable operators to manage network pressures while maintaining high service quality.

The primary goals for network investments in AI

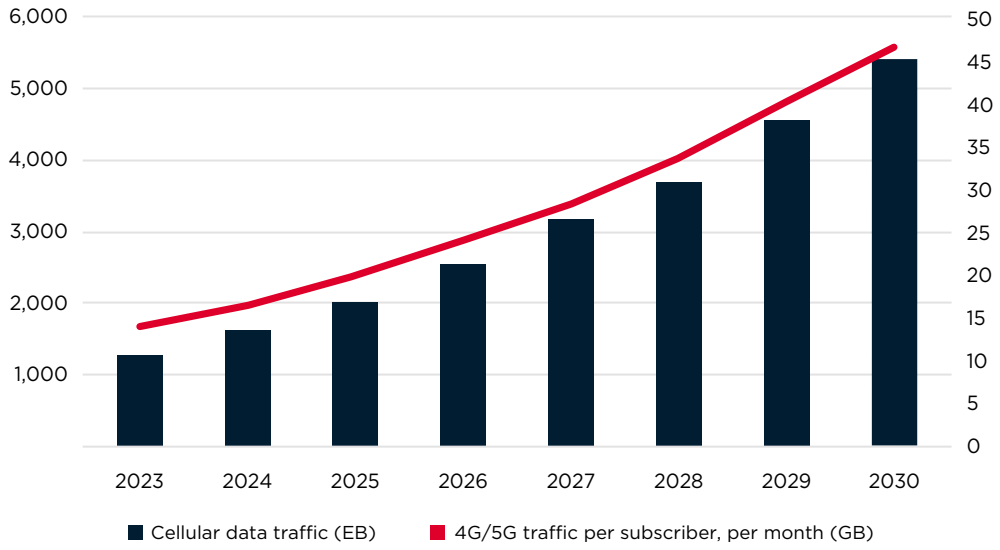
	1	2	3
2024	Efficient network operations	Improved customer experience	Revenue generation
2026	Improved customer experience	Efficient network operations	Revenue generation

Source: GSMA Intelligence Network Transformation Survey 2024

Mitigating insatiable rises in traffic...

- **Traffic surges.** There is significant traffic growth from consumer (e.g. ChatGPT) and enterprise applications (e.g. inferencing). Cellular data traffic is expected to triple by the end of the decade, largely due to 5G adoption, with the added load from genAI pushing traffic demand even higher. While this poses significant challenges for network operators, AI-driven traffic management offers a solution to optimise performance and handle the increased load.
- **The role of genAI.** AI technologies such as traffic routing and load balancing can intelligently distribute network traffic, reducing congestion and improving user experience. For example, T-Mobile, in partnership with OpenAI and Nvidia, developed a self-optimising network using AI to dynamically route traffic and balance loads. In addition, predictive maintenance with AI allows operators to detect and address network issues. Ericsson, for example, offers an AI solution to predict and resolve anomalies before they impact customer experience, ensuring greater network reliability amid growing traffic demand.
- **Bandwidth demand grows.** While AI offers solutions to mitigate rising traffic, bandwidth demand will grow. Operators will need further infrastructure upgrades. Further, the surge impacts RAN, core, backhaul and OSS/BSS, necessitating both efficiency improvements and increased capacity to handle growing data loads.

Data traffic growth may be slowing, but it will still grow almost threefold by 2030

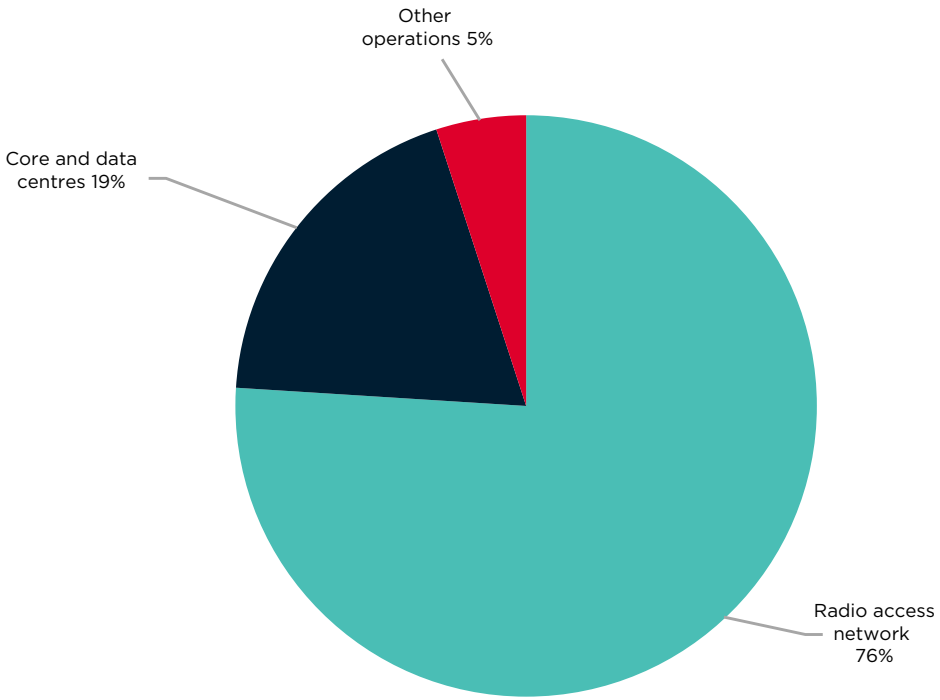


Source: GSMA Intelligence

...and stubborn energy usage

- **RAN on top.** For operators, the RAN accounts for the highest investment levels. It also accounts for almost three quarters of the total energy consumption. Energy remains a stubborn cost line, accounting for 20–30% of network opex.
- **Tackling the challenges.** As networks expand, AI-driven energy optimisation will be critical to balance sustainability goals with increasing traffic demand. The energy impact with data centres is significant, accounting for around 20% of total energy consumption. With rising bandwidth requirements and high energy use, demand for AI in network operations is growing. However, several challenges must be addressed:
  - **Seamless interoperability** – AI solutions must integrate across diverse data sources within complex network infrastructures.
  - **Scalability across nodes and devices** – As networks grow, AI-driven solutions need to scale efficiently, but this remains a significant challenge.

The RAN accounts for three quarters of operators' energy draw



Source: GSMA Intelligence Telco Energy Benchmark 2024

Having a smarter network and working together

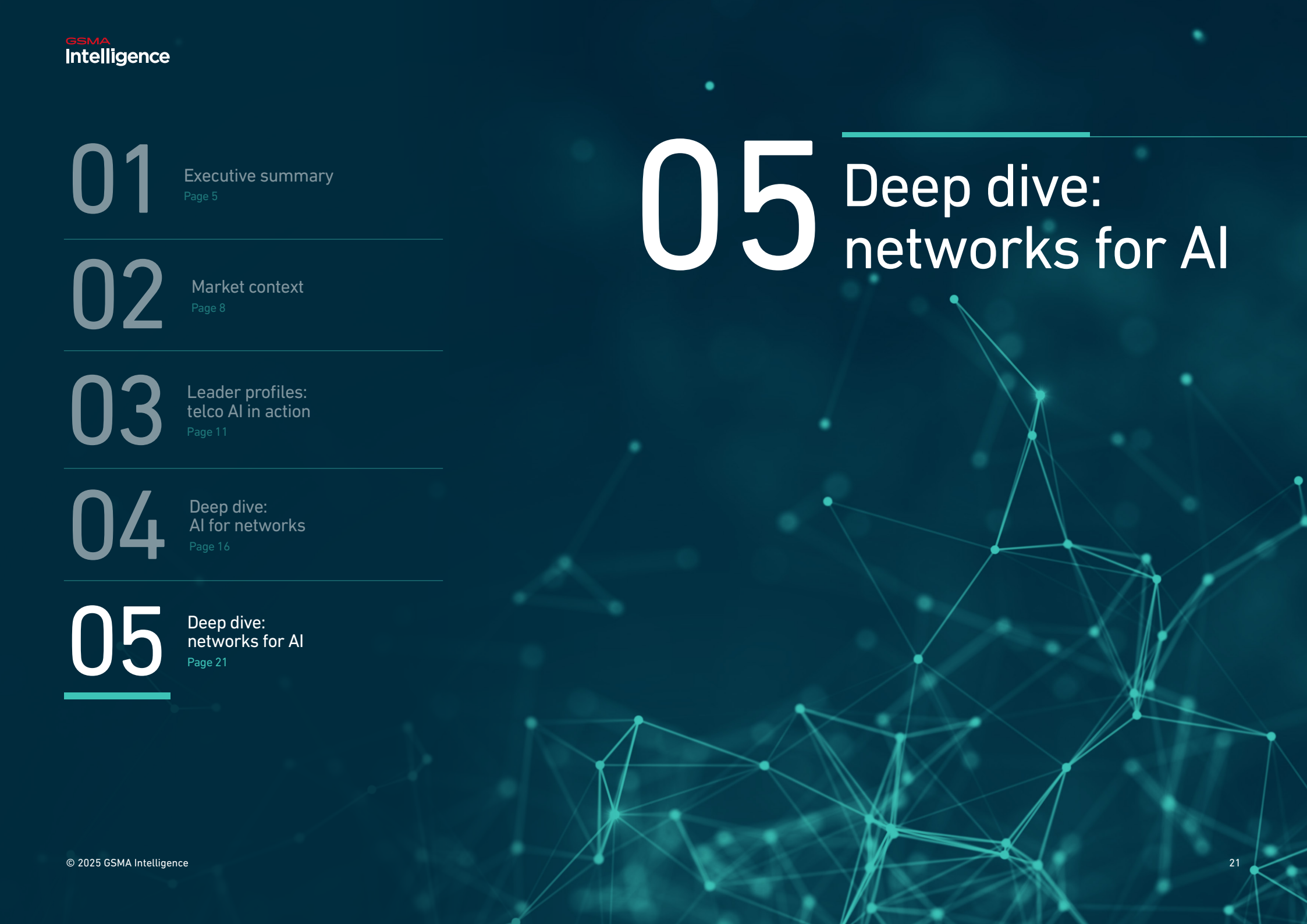
- **AI RAN opportunities.** Operators leveraging AI for the RAN can achieve significant benefits. Since the RAN is one of the most expensive and energy-intensive components of a network, AI-driven optimisation can lead to significant cost savings and performance improvements.
- **How AI is helping.** AI-driven traffic steering and RIC can help optimise resource allocation and improve network efficiency. AI automates network configuration, fault detection and anomaly resolution, minimising manual interventions and enabling real-time decision-making. By leveraging AI, operators can achieve greater efficiency, lower costs and enhanced network performance, ensuring scalability for future traffic demand. Operators such as SoftBank and T-Mobile are moving into AI RAN innovation, focusing on the convergence of AI and RAN architectures to improve network performance.
- **Taking steps together.** To make the most of AI, industry-wide efforts are crucial to help build common frameworks and best practices. Initiatives such as the AI-RAN Alliance, with operators and technology firms, aim to develop shared approaches for AI integration in the RAN. Another such initiative is O-RAN Alliance, which is working towards standardisation on AI-driven RAN initiatives. Such alliances can also better address issues such as interoperability and scalability, and help fast-track innovation.

Examples of AI RAN initiatives from the ecosystem

RAN initiative	Led by	The plan
AI-RAN Innovation Centre	T-Mobile US, Nvidia, Ericsson, Nokia	The centre aims to unite innovation in RAN and AI to deliver transformational network experiences for customers.
Native-AI RAN computing	Smartfren, ZTE	RAN computing integrates communication and computing directly inside the base station, unleashing the full potential of network infrastructure. Positioned as the network element closest to users, RAN computing achieves efficiency by precisely allocating radio resources according to service requirements and device capabilities. This can improve user experience by 15% and increases network capacity by 5%.
Nvidia AI Aerial	Nvidia	Nvidia AI Aerial is a suite of accelerated computing software and hardware for designing, simulating, training and deploying AI RAN technology for wireless networks in the AI era. It supports a growing ecosystem for AI RAN.
AI network planning	Telefónica, Google Cloud	The initiative involves testing an AI-based network planning tool, including planning live network upgrades and expansions. The new system delivers “twice as many accurate predictions” as previous network planning tools.

GSMA Intelligence

05 Deep dive: networks for AI



Enterprise workloads are moving to cloud and edge

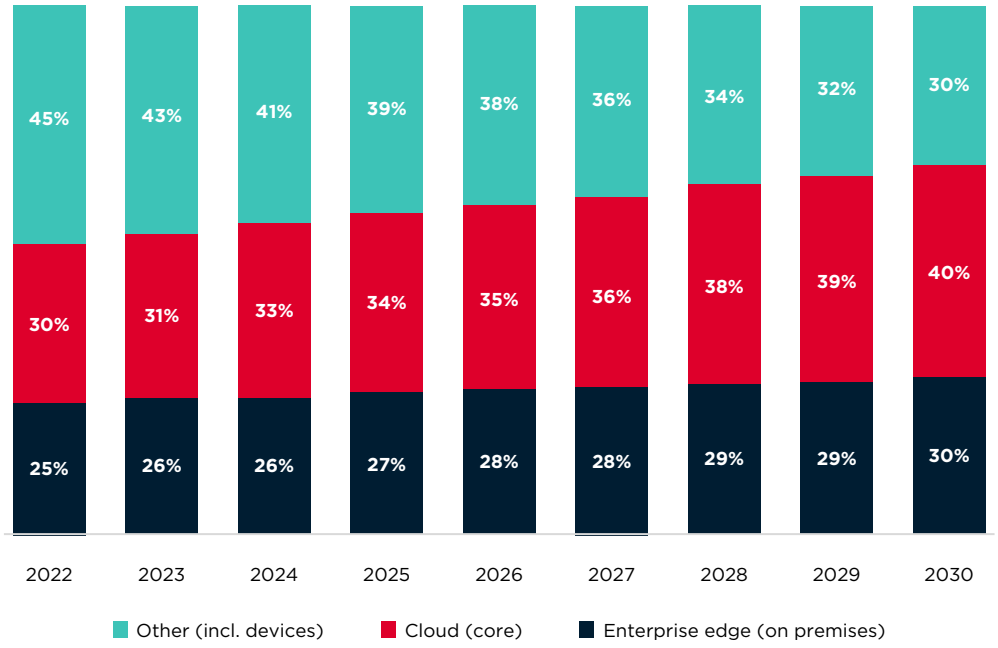
Networks for AI. Having AI-enabled, smarter networks is only part of the equation. Operators need stronger networks to support AI applications. As AI itself requires significant processing power, cloud and edge computing are essential

Cloud integration's role. Cloud integration enables operators to handle the massive data essential for AI workloads. Cloud data centres have the largest capacity to handle compute workloads, so are central to LLM training. However, compute power has progressively moved out along the spectrum, with 5G a catalyst for mobile operators selling into enterprise clients that want lower latency services closer to operations.

The importance of edge computing. Retaining data at the edge, as opposed to sending it to cloud data centres, could save energy and make money – a double benefit. However, AI adds a new dimension to the value of edge through distributed inference, with several attributes now supporting the strategic rationale for operators. These include:

- effectively managing and distributing compute processing loads
- data sovereignty and security
- improving energy efficiency
- resilience
- financial benefits: cost savings and new revenue streams.

Share of enterprise data traffic processing

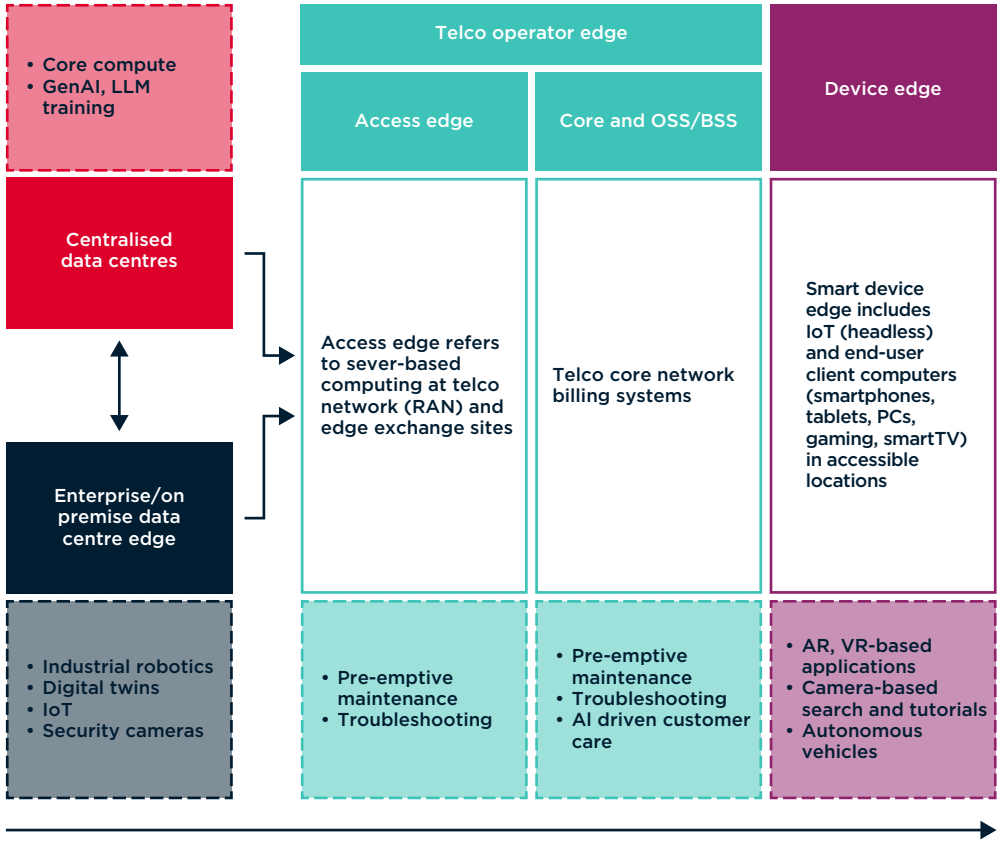


Source: GSMA Intelligence

AI inference brings new use cases to bear

- **Cloud, edge and inferencing.** Compute power becomes increasingly decentralised the further it moves from core data centres (such as those hosted by AWS). This allows operators to allocate resources efficiently where they are needed most. In cloud computing, data is processed in centralised servers. In edge computing, it is processed closer to where it is generated, on devices such as sensors or local servers at the edge of the network. AI inference is a process where a trained AI model can make decisions based on new data.
- **AI inference at the edge.** AI inference has a role in many parts of telecoms network structure. Where inference adds value comes down to the nature of the service – its compute and bandwidth requirements, and customer benefits. This enables various use cases and applications across different points in the network and offers flexibility, which is crucial to optimise performance, security and efficiency in diverse scenarios.
- **Diverse use cases.** The use cases start at the enterprise edge and move progressively out to the device edge, which could be a consumer device such as a smartphone, or a different form factor such as a car in the case of autonomous vehicles.
    - With enterprise edge, the primary objective is revenue generation.
    - With the telco far edge, at the RAN level, the primary objectives are cost savings and reducing operational risk.
    - At the telco near edge, the primary objectives are cost savings and reducing operational risk.

Mapping AI use across the mobile network

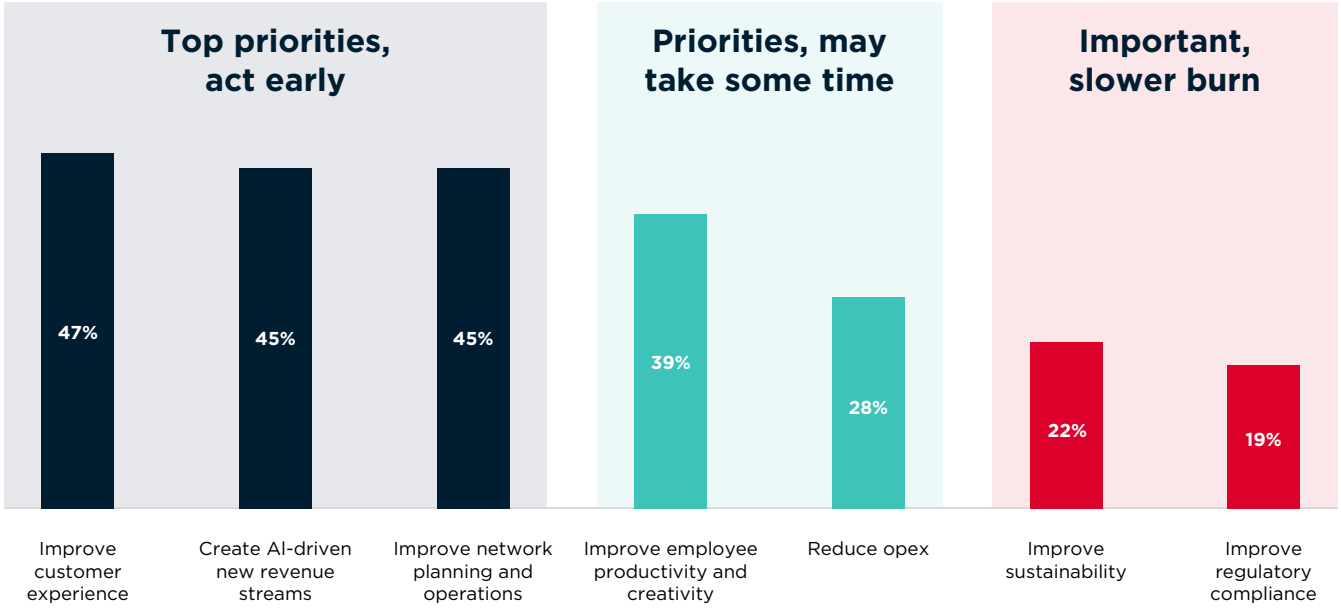


Source: GSMA Intelligence

Driving efficiency, revenue and customer value

- **Assessing the rationale for investing in AI.** Data from a GSMA Intelligence AI survey reveals the range of operator priorities for using AI:
  - priorities to get right early on (e.g. network optimisation, making money)
  - things that are important but may take time (perhaps due to the time to deploy or harvest benefits)
  - slower burns (e.g. energy consumption).

AI is about efficiencies, but – more than that – it's about revenue and customers
What are the three most important goals your company expects to achieve using AI?



Source: GSMA Intelligence AI benchmark based on survey of operators (N=100, H2 2024)

A range of AI revenue models are emerging

- **Making money from AI.** Nearly 50% of operators rate revenue as a priority from AI. Much of the early AI experimentation reflects this sentiment. Examples include the following:
  - Selling GPU as a service – operators can incorporate GPUs from silicon players and offer AI processing through their own compute and data centre resources.
  - Sell-through models – these could include selling proprietary data for LLM training in non-English languages, AI-enabled services for enterprises or consumer-focused propositions.
  - AI factories – AI compute, network capabilities and AI platforms are offered together as comprehensive services. Telenor's AI factory is an example.
- **Act early.** The options above are far from mutually exclusive, and many operators may offer services in several categories. The imperative is to act early as there is a competitive window

Examples of revenue opportunities with AI

Operator and AI provider	Investment	Coverage	Revenue model
Vodafone and Google	$1 billion over 10 years	Europe, Africa	AI connectivity partner + services: B2C (Pixel distribution, pay-TV upsell, security)
KT and Microsoft	'Multi billion dollar', including $450 million KT fee	South Korea	AI compute provider – sovereign cloud: B2B (LLMs for core business and enterprise services)
Orange and Meta, Open AI	Not reported	Europe, Africa	AI solutions partner – model development: B2B (language translation for customer care), non-profit
KDDI and AWS	Not reported	Japan	AI solutions partner – consultancy services: B2B (LLM and app support for enterprises, government)

Source: GSMA Intelligence

Managing network requirements with AI

- **Network readiness.** The impending traffic rise from AI (with an uplift by 2030 of potentially 50% on what traffic would otherwise be) means there is a need for effective distribution of workloads to the edge to balance out processing in the central cloud. AI builds on automation solutions already available, though we expect more innovation in LLM development specific to mobile networks, so that enhanced features such as dynamic traffic shunting can be implemented.
- **The role of network slicing.** While cloud and edge computing and 5G-ready networks will be essential in ensuring that infrastructure can support the increased volume of data and processing needs of AI applications, network slicing will be critical to provide the required quality of service for different types of AI workloads. AI applications, such as real-time inferencing, deep learning models and autonomous systems, require different levels of latency, bandwidth and compute power. Slicing ensures customised, scalable and efficient connectivity in industrial automation and emergency services, as well as for large-scale public events, for example.
- **Downstream impact.** There are also impacts for other parameters such as energy consumption and costs. Distributed compute again plays a key role by retaining data for processing at the edge, as opposed to sending it to the cloud. This offers energy savings by avoiding the power associated with backhaul transport and data centre cooling. The savings are in the order of 50–60%. For example, by retaining 30% of traffic at the edge rather than sending it to the cloud, an enterprise could save around 20% on energy usage and overall total cost of ownership.