# Energy Implications of AI for Mobile Operators

April 2026

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 **gsma.com**

GSMA Intelligence is the definitive source of global mobile operator data, analysis and forecasts, and publisher of authoritative industry reports and research. The data we generate and the analysis we do 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.

**gsmaintelligence.com**
**info@gsmaintelligence.com**

Ernst & Young (EY) is building a better working world by creating new value for clients, people, society and the planet, while building trust in capital markets.

Enabled by data, AI and advanced technology, EY teams help clients shape the future with confidence and develop answers for the most pressing issues of today and tomorrow.

EY teams work across a full spectrum of services in assurance, consulting, tax, strategy and transactions. Fuelled by sector insights, a globally connected, multidisciplinary network and diverse ecosystem partners, EY teams can provide services in more than 150 countries and territories.

All in to shape the future with confidence.

EY refers to the global organisation and may refer to one or more of the member firms of Ernst & Young Global Limited, each of which is a separate legal entity. Ernst & Young Global Limited, a UK company limited by guarantee, does not provide services to clients. Information about how EY collects and uses personal data and a description of the rights individuals have under data protection legislation are available via **ey.com/privacy**. EY member firms do not practice law where prohibited by local laws. For more information about our organisation, please visit **ey.com**.

# Contents

# Executive summary

**Artificial intelligence (AI) is transforming telecommunications and raising new energy questions for the mobile industry.** As AI becomes more widely adopted and embedded across the value chain, it is critical for telco operators to understand the potential energy implications of AI – both from the growing energy demands of AI in data centres and networks, as well as its potential to accelerate network energy efficiency. This report examines the direct and indirect energy impacts of AI on the telco sector, providing an analysis of current trends and future projections to help operators navigate this transformation.

**AI is driving significant growth in data centre energy demand, but only certain regions and operators will feel the impacts in the near term.** Driven by the rapid expansion of AI data centres, global data centre electricity use is projected to more than double between 2024 and 2030. Near-term growth is expected to be concentrated in hyperscale cloud facilities that host generative AI (GenAI) workloads for model training and inference. Data centres owned by telecommunications companies (telcos) account for only one-tenth of total data centre energy use today and have a far more modest growth trajectory. However, regional impacts vary considerably, with some telco operators in Asia, the Middle East and Africa likely to see rapid growth in energy consumption as they expand their AI data centre capacity in electricity grids that are heavily dependent on fossil fuels. In contrast, telco operators in Europe and North America are generally less exposed because they typically do not operate data centres that handle heavy AI workloads.

**Early indications are that direct and indirect AI-related traffic could account for 15–60% of mobile data traffic by 2030, but AI-related traffic is not expected to be a major driver of network energy demand growth.** GenAI applications represent less than 1% of total data traffic today, with a high degree of uncertainty in future growth. Direct AI-related traffic is projected to account for 5–20% of total mobile data traffic by 2030, with a further 10–40% from indirect AI-related traffic.  However, because energy use in both fixed and mobile access networks is driven primarily by capacity, coverage and hardware configuration, rather than proportionally to traffic levels, AI-related traffic is not expected to be a primary driver of future network energy demand over the next five years. This means that the direct impact of GenAI on network energy is expected to remain modest in the short term, with the main risk coming from long-term capacity expansion.

**AI solutions have the potential to reduce current network energy consumption by up to 30%, but deployment and scaling are difficult.** AI-driven dynamic power management, predictive maintenance, traffic optimisation and intelligent data centre cooling are delivering 5–15% energy savings in live deployments, with related applications also helping to extend equipment life and reduce overall energy consumption and emissions. In some cases, AI-enabled solutions have achieved up to 40% reductions in total energy consumption in the radio access network (RAN) power or cooling systems. However, most operators are still in early stages, with pilot projects delivering strong local results but limited system-wide adoption. Scaling these solutions from pilots to production will require building trust, interoperability and robust change-control processes.

**The future energy impacts of AI on mobile operators are uncertain and will depend on where AI workloads are processed, how fast AI adoption grows and how quickly operators develop the capabilities to manage it.** In the short term, the impact of AI on network energy is expected to be modest. By the late 2020s, rising inference workloads and AI-driven traffic could shift growth in energy demand from networks to telco-owned computing infrastructure, particularly for operators investing in their own data centres. Over the long term, AI could radically change how mobile networks interact with the energy grid, requiring grid-aware scheduling and participation in energy flexibility markets, much like some operators are trialling today.

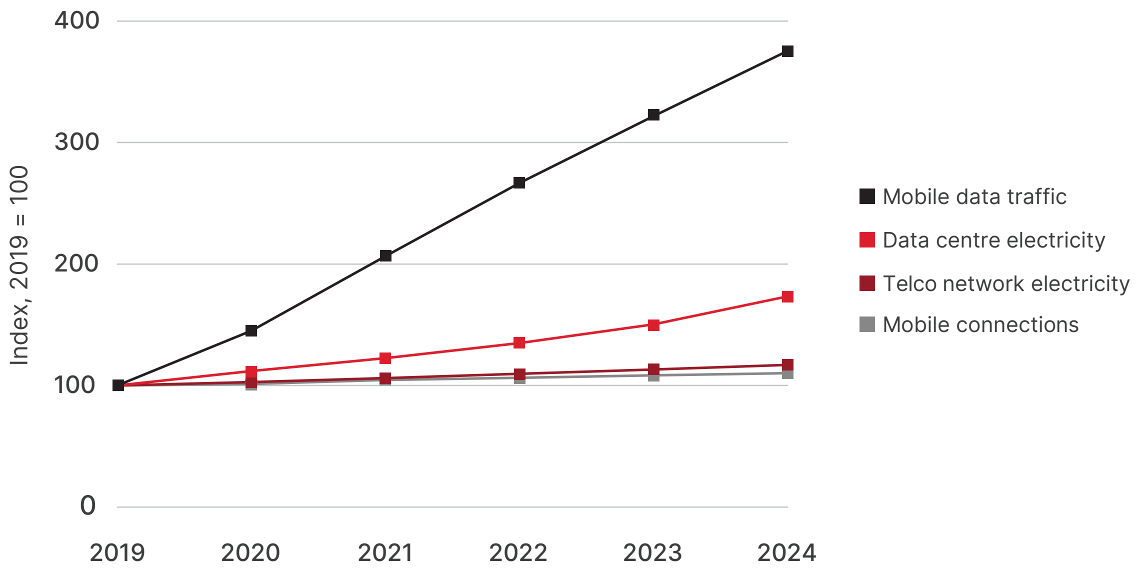
**The overall impact of AI on mobile operators' energy and sustainability objectives will largely depend on how AI is implemented in practice.** Operators that invest early in measurement, governance and interoperability will be better positioned to capture AI-driven efficiency gains and manage the additional energy demands that may arise as AI use scales. For operators building and operating AI data centres, sourcing renewable energy is a critical lever to manage long-term costs and support net-zero commitments. At the same time, broader outcomes will also depend on supportive policy frameworks, particularly around renewable energy access and grid modernisation, where collaboration between operators, suppliers and governments will be essential.

# 01 AI and energy: A double-edged sword

The past decade has seen an extraordinary expansion in digital connectivity. More than 5.5 billion people, two-thirds of the global population, are now online, with mobile their primary route for accessing the internet.[1] There are nearly 9 billion cellular connections and around 7 billion smartphones worldwide – numbers that have tripled in a decade.[2] Global internet traffic has grown 25-fold since 2010 and is expected to triple again by 2028.[3]

Despite this soaring demand, the energy use of telco networks has risen only modestly due to strong efforts in energy efficiency. In 2023, consumption was approximately 290 TWh (around 1% of global electricity use), up just 12% since 2019.[4] Nonetheless, energy remains a significant cost for operators, typically reaching up to 15-30% of network opex, making ongoing efficiency improvements both an environmental and a financial priority for the telco sector.[5]

**Figure 1** | Growth of data traffic, connections and electricity consumption, 2019-2024




Sources: IEA, GSMA and Ericsson.

It is uncertain whether gains in energy efficiency can continue to keep pace with growing demand for energy, particularly as AI expands. Telecommunications companies (telcos) have been among the early adopters of AI, using techniques such as machine learning (ML) to optimise networks, analyse service disruptions, predict maintenance needs, enhance customer call-centre experience and enable new digital services. Simultaneously, telco operators provide the connectivity that enables others to use AI across the global economy, and some are also building AI data centres.

Investment in AI and data centres is booming, raising concerns about the growing energy and environmental footprint of AI. Global investments in data centres grew 51% year-on-year in 2024 to $455 billion, with around $170 billion of new assets required to be constructed in 2025 alone.[6] By 2030, data centres are estimated to need nearly $7 trillion in global investment to meet surging demand.[7] Consequently, data centre energy consumption is anticipated to more than double by the same year, driven by a rise in AI-related workloads. In some areas, such as Ireland and the US state of Virginia, data centres already consume more than a fifth of total electricity.[8, 9]

1 IEA (2025) Energy and AI.
2 Ericsson (2025) Ericsson Mobility Report June 2025.
3 EY and Liberty Global (2025) Smarter Networks, Greener Planet.
4 GSMA (2025) Mobile Net Zero 2025: State of the Industry on Climate Action.
5 GSMA analysis based on GSMA Intelligence (2026) Going green: measuring the energy efficiency of mobile networks and towers.
6 Ashare, M (20 March 2025) Data center investments surged to $455B last year: report. CIO Dive; Craske, B (13 October 2025) Financing the Future: Trends in 2025 Data Centre Investment. DataCentre Magazine.
7 Law, M (14 May 2025) AI Infrastructure to Require $7tn by 2030, says McKinsey. DataCentre Magazine.
8 EPRI (2024) Powering Intelligence: Analyzing Artificial Intelligence and Data Center Energy Consumption.
9 Central Statistics Office (10 June 2025) Data Centres Metered Electricity Consumption 2024.

## What is artificial intelligence?

Artificial intelligence refers to computer systems designed to perform tasks that typically require human intelligence, such as learning from experience, recognising patterns, making decisions, understanding natural language and solving complex problems. AI systems use algorithms and computational models to process data, identify relationships and generate outputs that would traditionally require human cognitive abilities.

Generative AI (GenAI) is a subset of AI that creates new content, such as text, images, code, audio or video, based on patterns learned from training data. Traditionally, AI systems were primarily designed for discriminative tasks, like forecasting future demand or categorising images, whereas GenAI is designed to produce new outputs, like text, speech, images or video that resemble the training data. This distinction is particularly relevant for energy, as GenAI typically requires substantially more computational resources and energy, both for training models and running them to respond to queries (known as inference) compared to conventional AI applications. This is due to their larger scale and the complexity of generating novel outputs. While training may require substantial amounts of energy for a relatively short time, inference is less energy intensive but can consume significantly more over the long term due to increases in AI use.


For telco operators, the increasing adoption and use of AI within the network, in data centres and in support of other operational activity represents a “double-edged sword”.

On the one hand, AI offers substantial new opportunities to make networks even more energy efficient, over and above the long-standing use of conventional AI by operators and other aligned efforts to modernise networks, such as transitioning from 2G/3G to 4G/5G and replacing copper cables with fibre. AI-enabled solutions can deliver 10–40% reductions in (RAN) power consumption, reduce cooling system requirements by 10–15% and enable predictive maintenance that extends the life of equipment and reduces embodied energy and carbon (for materials and manufacturing) from replacements.[10]

10 Ernst & Young LLP (2025) Smarter networks, greener planet.

On the other hand, increasing consumer and business use of AI could reshape data traffic patterns in significant ways. Analysis from GSMA Intelligence and Ericsson shows that AI applications exhibit fundamentally different network behaviour compared to traditional mobile services, including:[11, 12]

**Video-based AI interactions**

Applications using real-time video feeds, such as AI assistants processing camera input and visual searches or virtual/extended reality, are particularly data intensive and require continuous uplink and downlink traffic, depending on codecs and how the AI is deployed. These applications drive substantially higher per-user data consumption.

**Uplink-dominated traffic**

While typical mobile traffic follows a 90/10 downlink-to-uplink ratio, GenAI applications have roughly a 75/25 ratio, nearly triple the normal rate of uplink. This is driven by users uploading images, videos and voice data for AI processing.[13]

**Indirect AI-generated content**

AI tools that make it easier to create video, images and audio for social media and streaming services may significantly increase traffic. Users can now generate professional-quality content without specialised skills.

**Agentic AI communications**

The much-anticipated growth in agent-to-agent communication as enterprises deploy AI to automate manual processes and data analysis may generate intense bursts of traffic. These autonomous applications could create unpredictable background traffic, complicating network capacity planning.

Collectively, these changes make future traffic volumes and patterns difficult to predict, with scenarios from GSMA Intelligence suggesting that combined direct and indirect AI-related traffic could reach 15–60% of total mobile data by 2030. It is also uncertain how these changes could impact network energy use, since the energy use of modern access networks is not proportional to data traffic volumes, even allowing for anticipated energy reductions from 6G compared with 5G baselines.[14, 15]

These issues raise key questions for operators, including:

- How can growth and innovation be balanced with energy efficiency and environmental responsibility?
- Where will the energy impacts of AI be felt most – in the network or in data centres?
- How will AI reshape traffic patterns, and what will this mean for network energy use over the next five to 10 years?
- How will other AI use cases increase energy consumption, both in the short term and in the future?

11 GSMA Intelligence (2025) Telco AI: State of the Market Q1 2025.
12 Ericsson (2025) Understanding the Impact of GenAI. Ericsson Mobility Report June 2025.
13 Ericsson (2025) Exploring GenAI's Impact in 2025. Ericsson Mobility Report June 2025.
14 Mytton, D., Lundén, D. and Malmodin, J. (2024) Network energy use not directly proportional to data volume: The power model approach for more reliable network energy consumption calculations. Journal of Industrial Ecology, 28(4), pp. 966–980.
15 Samsung (2025) Energy Saving for 6G Network: Part II, From Always-ON to Smart-ON.

This report analyses the energy implications of AI in the telco sector. Throughout the report, we distinguish between **direct energy impacts** (for example, the electricity used by **telecom**munications networks and any telco-operated data centres) and **indirect or systemic impacts** (for example, how AI-enabled services change demand and energy use elsewhere in the economy). The following sections cover the direct and indirect energy impacts of AI on telcos, the opportunities to use both more traditional AI techniques and GenAI to improve energy efficiency and the risks, future trends and regulatory landscape.

The report draws on modelling from GSMA Intelligence, interviews with telco operators and suppliers and industry research and case studies, to provide decision-makers with evidence-based guidance on responsible and sustainable AI adoption.

# 02 Understanding the energy impacts of AI

In the past 10 years, network operators have captured a "first wave" of energy-efficiency gains by modernising networks and infrastructure. These strategies, described in **Table 1**, do not use AI but have allowed operators to carry rapidly rising data volumes with flat or declining energy consumption per connection.

**Table 1** | Traditional (non-AI) energy-efficiency measures[16, 17]

| Approach | Description | Typical efficiency gains |
|---|---|---|
| Passive infrastructure upgrades | Replacing copper with fibre, modernising power supplies, efficient hardware | 10–30% |
| Network modernisation | Upgrading from 2G/3G to 4G/5G, sunsetting legacy networks and increasing use of network virtualisation | 15–40% |
| Site simplification | Consolidating sites, reducing redundant equipment | 10–20% |
| Scheduled power management | Manual or timer-based shutdowns during low-traffic periods | 5–10% |

Fibre-based networks consistently show much lower electricity consumption compared with fixed wireless and legacy copper technologies.[18] For example, Europacable analysis suggests that fibre-to-the-home (FTTH) gigabit passive optical networks (GPON) consume less than a third of the energy per subscriber than fixed wireless access – ignoring the contribution from customer premises equipment.[19] Liberty Global benchmarking similarly finds that fibre roll-outs can deliver up to an 85% improvement in fixed network efficiency, while retiring 2G/3G networks typically reduces operator electricity use by around 15%.[20] Due in part to these modernisation efforts, electricity use per connection has been flat since 2019 (around 30 kWh), despite a fourfold increase in data traffic.[21]

While methods such as virtualisation and site simplification remain crucial, continued gains may become more difficult due to the high cost of infrastructure upgrades, the complexity of integrating new technologies with legacy systems and the need for new skills. Research carried out by GSMA Intelligence suggests that a quarter (25%) of operators now believe that the most effective way to improve energy efficiency in active infrastructure is AI and resource optimisation.[22]

The integration of AI in telecommunications is reshaping both the operational landscape and the energy profile of the sector. The impact of AI on energy consumption can be grouped into two categories, as illustrated in **Table 2**.

16 EY and Liberty Global (2025) Smarter Networks, Greener Planet.
17 GeSI (2024) Digital Infrastructure Modernization for an AI Future: Opportunities and Challenges Ahead.
18 Ibid.
19 Ibid.
20 EY and Liberty Global (2025) Smarter Networks, Greener Planet.
21 Ibid.
22 GSMA Intelligence (2024) Going Green: Measuring the Energy Efficiency of Mobile Networks.

**Table 2** | Examples of direct and indirect energy impacts of AI on telco data centres, networks and operations

| Impacts | On data centres | On networks | On operations |
|---|---|---|---|
| **Direct impacts** in the form of energy demand from AI workloads in data centres and the traffic they generate | Increased computing power for AI training and inference<br><br>Higher cooling loads from denser AI chips<br><br>Growth of telco-owned and edge AI data centres | Additional traffic from AI-based services<br><br>Higher data processing needs in access and transport networks<br><br>Potential long-term capacity expansion | Greater cloud and IT use for internal AI tools<br><br>Added compute for analytics and customer-facing AI applications |
| **Indirect impacts** in the form of energy savings from the use of AI for network optimisation and other operational activities | Energy savings from optimising cooling<br><br>Energy savings from smarter workload scheduling<br><br>Use of AI to predict server failures and reduce downtime | Energy savings from AI-driven network sleep modes<br><br>Energy savings from traffic and signal optimisation<br><br>Reduced energy waste through dynamic power management | Energy savings from predictive maintenance and fewer site visits<br><br>Improved network planning and asset use<br><br>Efficiency gains in logistics and field operations |

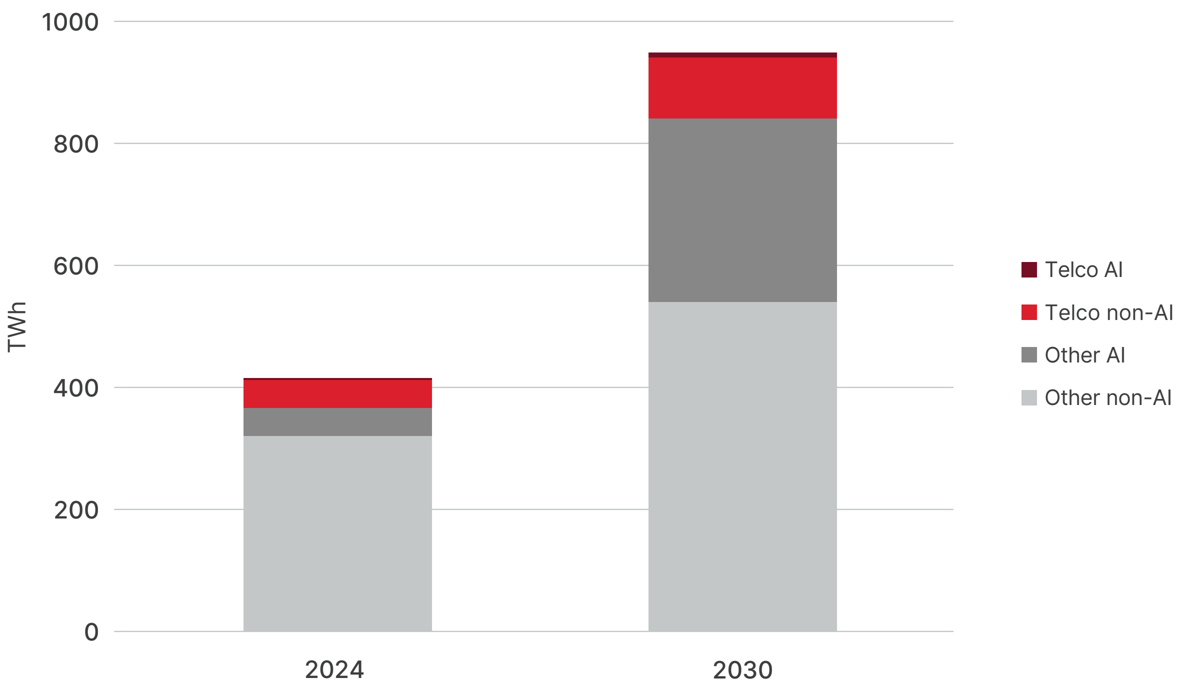
## 2.1 Direct energy impacts

### 2.1.1 Data centres

The most immediate energy impact of AI is the data centres where AI models are trained and run, also known as "AI workloads". After more than a decade of relative stability, the International Energy Agency (IEA) now projects a sharp increase in electricity demand, which is expected to rise from around 415 TWh in 2024 to nearly 945 TWh by 2030, or more than four times faster than global electricity demand.[23] AI workloads are expected to more than triple between 2025 and 2030, and will account for nearly half this increase.[24]

**Figure 2** highlights this shift, breaking down global data centre demand by workload type. While AI workloads are expanding rapidly, telco-owned facilities account for a very small proportion. In 2024, telco operator data centres consumed around 48 TWh of electricity, or around one-tenth of overall data centre energy consumption, as illustrated in **Figure 2**.[25, 26] This suggests that most large-scale AI training and inference is taking place outside the telco sector.

**Figure 2** | Global energy demand by data centre ownership and type of workload




Note "Other" are all non-telco owned data centres, including hyperscale, colocation and service providers, and enterprise data centres not owned by telcos. Accelerated server energy consumption is used as proxy for AI energy use.

Source: Telco share based on GSMA Intelligence; global totals from IEA. (2025). Energy and AI.

Operators are more directly exposed in inference workloads and AI-native applications that run at the edge. For example, a single AI-powered search or recommendation can require more than 10 times more electricity than a basic Google search.[27] While volumes are still relatively small, more widespread use of AI-powered search and other types of inference could drive up energy consumption. Some operators in Asia are experimenting with "on-premises" large language model (LLM) inference, such as Chunghwa Telecom in Taiwan.[28]

23 IEA (2025) Energy and AI: World Energy Outlook Special Report.
24 McKinsey & Company (1 August 2025) Scaling bigger, faster, cheaper data centers with smarter designs.
25 Ibid.
26 IEA (2025) Energy demand from AI.
27 O'Donnell, J. and Crownhart, C. (20 May 2025) We did the math on AI's energy footprint. Here's the story you haven't heard. MIT Technology Review.
28 GSMA (2024) Operator Best Practices: AI Large Model Empowering Verticals Use Cases.

## Understanding data centre types

| | Type of data centres | | | | |
|---|---|---|---|---|---|
| | Edge | Enterprise | Colocation and service provider | Hyperscale | |
| | | | | General purpose | AI-focused |
| **Definition** | Small, distributed facilities located close to end users to minimise latency; often deployed in telco facilities | Facilities owned and operated by companies for their own internal computing needs | Facilities that lease space to customers (colocation) or provide both space and computing resources (service providers) | Massive facilities operated by major technology companies, designed for extreme scalability and efficiency | Hyperscale facilities with combinations of CPUs and GPUs specifically designed for AI training and inference workloads |
| **Typical size** | 0.1–1 MW (micro and modular designs are common) | 1–10 MW | 10–50 MW | 50–300 MW (largest can exceed 1 GW) | 100–500 MW (some planned facilities exceed 1 GW) |
| **Example operators** | Telcos: AT&T, Deutsche, Telekom, Verizon, Vodafone | Corporate IT departments, financial institutions, healthcare systems | Cyrus One, Digital Realty, Equinix<br>Telcos: China Mobile, China Telecom, KDDI, NTTDATA, Telehouse | Alibaba, Amazon Web Services, Google, Meta, Microsoft Azure | OpenAI, Oracle, xAI |
| **Typical workloads** | 5G and core RAN functions, content caching, real-time analytics, augmented/ virtual reality, IoT and autonomous vehicles | Business applications, internal databases, email servers | Hosted applications, disaster recovery, cloud services for multiple tenants | Cloud services, web hosting, social media, AI training and inference, content delivery | AI training and inference |
| **Power usage effectiveness (PUE)** | 1.3–1.8 (higher variability due to smaller scale, diverse environments and cooling constraints) | 1.8–2.0 (less efficient due to age and scale) | 1.4–1.6 | 1.15–1.25 (most efficient due to advanced cooling, management) | 1.15–1.3 (advanced cooling needed for high-density compute) |
| **Share of global capacity** | Less than 5% currently, but growing rapidly with 5G roll-out and adoption of edge computing | Around 28% (declining from 85% in 2005) | Around 36% | Around 37% (up from 10% in 2010) | |

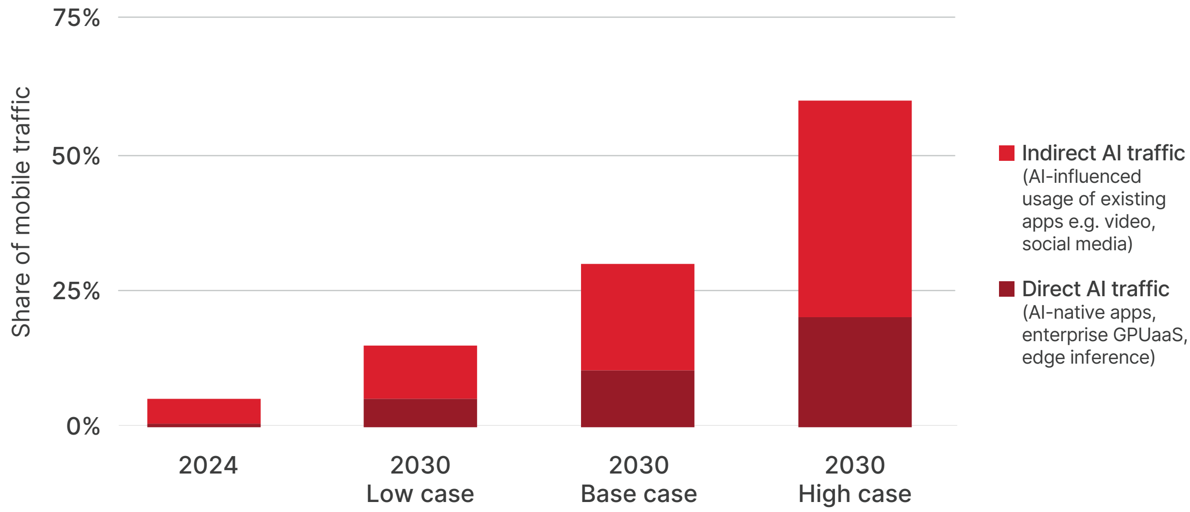
Note that telcos operate multiple types of data centres, including colocation facilities, edge facilities for 5G and low-latency services and internal network data centres for their own operational systems.

Source: Adapted from IEA (2025) Table 4.1, Energy and AI Report; LBNL (2024) United States Data Centre Energy Usage Report.

## 2.1.2 Data transmission networks

Based on data collected from operators, AI-related traffic appears to be negligible, accounting for around 0.1% of total mobile data.[29] However, modelling by GSMA Intelligence suggests a wide range of possible outcomes by 2030. For example, direct AI traffic could reach 5–20% of total traffic, while indirect effects, such as AI being used to generate video and audio, and greater use of AI in social media, could add another 10–40%.[30] These effects are illustrated in **Figure 3**.

**Figure 3** | Direct and indirect AI-related traffic as a share of total mobile traffic (2024 vs 2030 scenarios)




Direct AI refers to AI-native applications (consumer and enterprise), while indirect AI refers to AI-driven increases in the use of existing apps, such as video and social media.

Source: GSMA Intelligence (2025).

### Defining direct and indirect AI traffic

**Direct AI traffic** is data transmission directly associated with AI-native applications, where AI is the core function. Examples include consumer AI applications like ChatGPT and enterprise AI applications like copilots and chatbots. These types of AI are typically uplink-heavy because the user sends data to the cloud for AI processing rather than having it processed on their device.

**Indirect AI traffic** is data transmission from applications where AI enhances the experience but is not the primary function. Examples include AI-enhanced content on social media, intelligent non-player characters in games and spelling or grammar corrections in word-processing software. Data traffic for these applications tends to follow traditional patterns, but volumes may increase due to the use of AI for enhancement and personalisation.

Source: GSMA Intelligence (2025) Telco AI: State of the Market, Q1, 2025; Ericsson (2025) Mobility Report.

29 Ericsson (2025) Ericsson Mobility Report June 2025.
30 GSMA Intelligence (2025) Telco AI: State of the Market Q1 2025.

The impacts of higher AI-induced traffic on network energy use are likely to be limited due to the energy dynamics of access networks. This is because contemporary 4G and 5G network infrastructure typically operate at relatively constant power levels to maintain coverage and capacity, with radio equipment and base stations consuming similar energy whether transmitting at high or low utilisation rates.

Empirical data and recent research show that mobile network energy use is not driven primarily by changes in data traffic, as mobile networks continue to consume energy regardless of traffic volumes.[31, 32, 33, 34, 35, 36] Other factors, such as hardware configuration, network capacity and coverage and variable traffic patterns, are typically more important, both in the short- and long-term.

Existing network infrastructure will draw constant power regardless of traffic fluctuations. In the long run, the increasing density of 5G networks and expansion driven by greater capacity needs, rather than instantaneous traffic levels, will determine energy consumption. Therefore, the overall impact of AI-driven traffic changes will likely have relatively limited impacts on mobile network energy use. Fixed access networks, whose energy use is even less sensitive to data traffic, will likely have no impact on energy use as a result of AI-related traffic.

## 2.2 Indirect energy impacts

Beyond direct electricity use, AI may also have several indirect impacts on network energy. For example, AI can be used to significantly improve the efficiency of access networks and telco operations. AI also increases integration complexity, potentially driving up demand for edge-computing facilities to deliver low-latency applications and services to users, and triggering earlier-than-anticipated network upgrades if increasing adoption of AI applications requires network capacity to increase beyond current designs.

31 GSMA (22 June 2020) COVID-19 Network Traffic Surge Isn't Impacting Environment Confirm Telecom Operators.
32 Meta (2025) Electricity Consumption of Videos Viewed on Meta's Platform.
33 Rouphael, R.B. et al (2023) The impact of networks in the greenhouse gas emissions of a major European CSP". In 2023 International Conference on Electrical, Computer and Energy Technologies (ICECET), pp. 1-6. IEEE.
34 Mytton, D., Lundén, D. and Malmodin, J. (2024) Network energy use not directly proportional to data volume: The power model approach for more reliable network energy consumption calculations. Journal of Industrial Ecology, 28(4), pp. 966-980.
35 Guennebaud, G. and Bugeau, A. (2024) Energy consumption of data transfer: Intensity indicators versus absolute estimates. Journal of Industrial Ecology, 28(4), pp. 996–1008.
36 Schien, D. et al. (2025) Causal allocation of fixed impacts in product systems: Assessing the effect of data demand on network energy consumption. Journal of Industrial Ecology, 29(5), pp. 1618-1631.

## 2.2.1 AI as an enabler of efficiency

AI introduces a new set of dynamic, adaptive capabilities that can significantly enhance energy performance, as illustrated in **Table 3**.

**Table 3** | AI-driven energy efficiency initiatives[37]

| AI application area | What it does | Typical efficiency gains | Interaction with other levers | Case study example / source |
|---|---|---|---|---|
| **Dynamic power management (RAN/core)** | Learns traffic patterns and adjusts radios, carriers and features in real time | 10–40% of total RAN power (note that live deployments are often 10–20% whereas trials tend to be higher) | Substitutes manual or timer-based shutdowns; complements network planning and traffic steering | Sunrise: around 10% network energy reduction after AI roll-out Tele2/Nokia: AI can achieve 2–5 times the savings of non-AI timers |
| **AI-optimised data centre cooling** | ML tunes set points, airflow, chillers, etc. to cut cooling waste | 10–15% of cooling electricity and OPEX savings | Largely complementary to network drivers | Virgin Media O2: 15% reduction in data centre cooling at a savings of approximately $1.3 million/year |
| **Predictive maintenance** | Detects slow-developing faults; schedules condition-based maintenance; reduces truck-rolls and extends asset life | Approximately 5–10% of operational energy consumption plus avoided embodied energy from replacements | Complementary to all other AI opportunities; reduces rebound from failures/ over-provisioning | Telefónica: programme for operations optimisation, including predictive maintenance AT&T: AI used to pre-empt cable damage/field interventions |
| **AI-driven network planning** | Forecasts demand; optimises site count and placement and spectrum and transport; right-sizes capacity | 5–15% of energy over the full life cycle compared with building from scratch | Complementary to dynamic power management (planning sets the baseline) | Telefónica "Internet para Todos": AI used for siting/ route optimisation, reducing redundant builds |
| **Traffic routing optimisation** | Steers traffic to least-energy paths/edges; balances load to avoid hot spots | 5–10% typical intensity gains; large programmes can be higher | Complementary to dynamic power management | Comcast: 40% reduction in electricity per byte between 2019 and 2023 via virtualisation and AI/ML |
| **AI-enabled circular economy** | AI for device and asset triage, refurbishment, parts-harvesting and reverse logistics | Data is emerging (material/embodied energy likely to be more than operations) | Complementary (because it acts on embodied emissions/energy, not operations); typically separate from operational savings | Sunrise "Flex Upgrade": device-as-a-service and refurbishment pathways |

37 EY and Liberty Global (2025) Smarter Networks, Greener Planet.

In addition, AI video compression can achieve 30–50% better compression ratios than traditional methods while maintaining visual quality.[38] AI can also predict streaming and enable intelligent caching, moving content closer to end users and reducing long-haul network traffic.[39]

These AI-driven optimisation approaches could partially offset traffic growth from AI applications. However, its widespread deployment brings some challenges, including the need for industry-wide standards, integration complexity and the computational overhead of running AI optimisation systems.

### 2.2.2 Rebound effects and systemic risks

The relationship between efficiency, cost and demand follows a well-established pattern in telecommunications. Historical data shows that operators have consistently reduced per-bit costs, with the average cost per gigabyte for mobile data falling by more than 90% over the past decade. These falling costs, in combination with advances in digital technologies, have fuelled explosive growth in data consumption. However, AI introduces several factors that could accelerate this trend, including:

- **New AI-native demand categories** that did not previously exist, such as GenAI applications and agentic services, which raise the baseline power draw
- **AI-influenced demand** that makes content creation dramatically easier and leads to more users generating more video and automated content and richer interactive experiences
- **Uplink-intensive traffic patterns** that shorten sleep windows and raise baseline network power draw, particularly on time-division duplex (TDD) spectrum, where uplink-downlink balance must be actively managed
- **Always-on AI agents** that generate persistent background traffic for 24/7 automation, monitoring and automated decision-making

While past efficiency improvements were offset by increased usage, the impact of AI may be more complex because it simultaneously reduces per-transaction costs but expands the types and volume of transactions that are economically viable. The net result could be higher energy consumption overall. Unless policy controls and efficiency measures keep pace, data growth from AI could drive up network energy demand even as AI delivers savings in certain applications.

GSMA Intelligence research suggests that direct AI-native traffic will be 5–20% and indirect AI-influenced traffic 10–40% of mobile demand by 2030. Combined, AI-related traffic could represent approximately 15–60% of mobile data by the end of the decade (although these estimates are intended to be directional, not a precise forecast).[40]

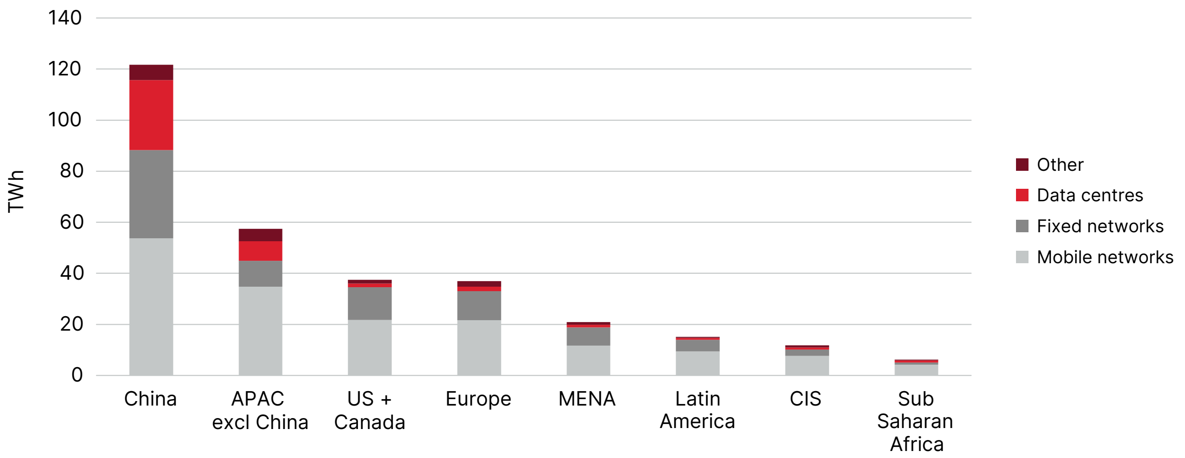
38 SuperAGI (30 June 2025) How AI is Transforming Video Encoding and Compression for Seamless Live Streams.
39 GCORE(3 September 2025) Smart caching and predictive streaming: the next generation of content delivery.
40 Data sourced by GSMA Intelligence (2025).

# 2.3 Operator energy landscape and trends

## 2.3.1 Networks vs data centres

For most telco operators, the radio access network (RAN) accounts for the majority of overall network energy use today.[41] Data centres typically account for less than 10% for most operators, but this figure varies by operator and region. However, industry reports suggest that this balance is expected to shift over time: data centres are likely to account for a growing share of operator energy consumption as AI inference becomes more deeply embedded in operations and customer-facing services.[42] Therefore, operators investing in AI-native data centres or hyperscaler partnerships will see this share rise.[43] This trend is illustrated in **Figure 4**, which compares estimated network and data centre energy use across regions in 2024. While network energy still dominates, the presence of substantial data centre consumption highlights the emerging shift in energy dynamics.

**Figure 4** | Energy use by telco operators, by region and end use (2024)




Source: GSMA Intelligence based on GSMA (2025) Mobile Net Zero 2025 and IEA (2025) Energy and AI.

[41] GSMA (2025) Mobile Net Zero 2025: State of the Industry on Climate Action.
[42] For example, see: McKinsey & Company (28 February 2025) AI infrastructure: a new growth avenue for telco operators.
[43] IEA (2025) Energy and AI: World Energy Outlook Special Report.

## 2.3.2 Regional variations in data centre energy impacts

Analysis by GSMA Intelligence suggests that the total energy used by operators and their corresponding energy growth profiles varies considerably by region of operation, as described in **Table 4**.

**Table 4** | AI-related data centre energy growth potential by region and type of data centre

| Region and energy growth potential | Overall data centre AI energy growth potential | Telco data centre AI energy growth potential |
|---|---|---|
| **China (High)** | ▶▶▶ **High impact**<br>Massive scale and increasing state-led AI development, with significant operator-owned data centre capacity | ▶▶▶ **High impact**<br>State-owned telcos (such as China Telecom) increasingly operate substantial data centre capacity, which is directly exposed to AI infrastructure demands[44] |
| **Asia-Pacific (excl. China) (High)** | ▶▶ **Moderate impact**<br>Mixed picture – policy and energy grid varies by market; Japan and Korea have significant operator-owned data centres | ▶▶▶ **Localised high impact**<br>Major exposure in Japan (for operators such as NTT and KDDI) and Korea (KT Corporation and SK Telecom) where telcos own or are investing in substantial facilities.[45, 46, 47, 48] In India, telcos are making substantial investments in data centre capacity (organisations such as Reliance Jio and Airtel)[49] |
| **United States (High)** | ▶▶▶ **High impact**<br>Largest absolute data centre growth and rapid AI infrastructure expansion amid continued grid interconnection challenges | ▶ **Limited impact**<br>Landscape dominated by hyperscalers; telco impact is primarily through edge/5G infrastructure rather than large-scale data centres |
| **Middle East and Africa (Moderate)** | ▶▶ **Moderate impact**<br>Lower current demand but steep growth trajectory; risk exposure is tied to grid capacity. Many operators in these markets still depend on diesel-powered generators, so AI will have limited impact until grids stabilise and renewable uptake increases | ▶▶ **Emerging moderate impact:**<br>Growing telco involvement in data centre development (e.g. Etisalat, Mobily, MTN); potential for increasing telco exposure as regional hubs develop |
| **Europe (Limited)** | ▶▶ **Moderate impact**<br>Strict regulatory environment and high energy prices leading to slower, hyperscaler-led data centre expansion | ▶ **Limited impact**<br>Telcos mainly building colocation and edge data centres (e.g. Deutsche Telekom and Telecom Italia);[50, 51] regulatory frameworks shifting the energy burden to hyperscalers |
| **Latin America (Limited)** | ▶▶ **Moderate impact**<br>Growing scale from a low base; risk is shaped by infrastructure constraints and availability of renewable energy | ▶ **Limited impact**<br>Small absolute scale; limited operator-owned hyperscale capacity |

Source: GSMA Intelligence.

44 Butler, G (5 July 2024) China Mobile launches data center in Beijing. DCD.
45 NTT Data (13 May 2025) NTT Data Expands Global Data Center Footprint with Land Acquisitions Across Seven Strategic Markets.
46 HPR (25 June 2025) KDDI and HPE join forces to launch AI data center operations by early 2026.
47 Butler, G. (26 March 2025) KT Cloud to establish AI data center demonstration center in Seoul, South Korea. DCD.
48 Ye-eun, J. (3 March 2025) SK Telecom unveils plan for Korea's largest AI data center at MWC 2025. The Korea Herald.
49 Communications Today (11 September 2025) Indian telcos redefine enterprise, connectivity and impact with AI.
50 Donegan, M. (8 August 2025) Deutsche Telekom lines up Brookfield for AI data centre investment. TelcoTitans.
51 Le Maistre, R. (3 October 2025) Telecom Italia targets B2B growth from sovereign services. Telecom TV.

The analysis shows that in regions such as China and Asia Pacific, and in other more localised hubs around the world, the potential for energy growth is shifting from the RAN to computing infrastructure as AI scales. This energy growth is being driven by a combination of state- and telco-led investments in data centre facilities, as operators seek to capture a share of the expanding AI-workload opportunity. However, in Europe, for example, the combination of regulatory constraints (and continuing uncertainty around new legislation, such as the Digital Networks Act) and high energy prices force operators to focus more closely on infrastructure that directly supports customers and services. Several operators have divested their data centres in recent years to focus on their core telecommunications business, turning to colocation data centre operators instead for infrastructure support.[52, 53] Other risk drivers, such as access to the electricity grid and the availability of energy, including renewables, creates a nuanced overall picture and continued uncertainty about how risk levels might change over time.

For telcos and mobile operators pursuing net-zero targets, the physical location of AI infrastructure has significant carbon implications. Running AI in regions where grids are mostly powered by renewables produces very different emissions than the same workloads running in fossil fuel-heavy grids. Scheduling AI processing adds another layer of complexity: inference jobs that can be delayed even by a few hours could run when wind and solar generation is at a maximum, rather than during traditional morning or evening demand peaks, which is often when fossil fuels are needed to bridge the gap.

However, without power purchase agreements (PPAs) to prove that new renewable capacity was built specifically to serve this demand, companies risk taking credit for renewable energy that already exists and would have been used anyway. Meanwhile, their new electricity demand is still being met by fossil fuel power plants, which typically fire up when grid demand increases.

The urgency of addressing the energy impacts of AI is due to its unpredictable growth, both in terms of volume and patterns of use, which threaten to erode the gains that telcos have made through equipment upgrades and more efficient cooling.

52 Menear, H. (26 August 2020) The complicated future of data centres and telco. Telco Magazine.
53 Hall, J. (4 March 2021) Telecoms on the edge: Is edge the answer to delivering 5G services?

# 03 Network efficiency and operator maturity

AI can either help networks reduce energy use or add new pressure, depending on the technology and the maturity of the operators deploying it. Interviews with telco operators and service providers confirm that adoption is uneven. Most are seeing early gains in pilots with a narrow scope, typically in RAN power management and data centre cooling, delivering efficiency savings in the 5–15% range. One European operator, for example, reported single-digit percentage network savings, while another based in Asia Pacific cited reductions of around 16% from its AI energy-management platform. Yet, these successes remain relatively isolated. Scaling them into durable, system-wide gains is difficult, with progress often constrained by legacy infrastructure, data quality problems and multi-vendor complexity. A second European operator described inconsistent asset naming and siloed telemetry as the biggest barrier to closed-loop automation, while another stressed that promising AI pilots struggle to reach production scale without rethinking governance and trust in AI systems.

Operators consistently reported that the RAN remains the largest consumer of energy, with data centre exposure climbing but at different rates across the industry. A European operator flagged energy grid congestion and location trade-offs as barriers to expanding its AI data centre capacity, while another, based in Asia Pacific, noted that quantifying the energy footprint of AI is “largely meaningless” given its evolving scope. Across the board, operators agreed that the net energy impact of AI is uncertain, as it could accelerate efficiency or raise baselines if not managed carefully.

These realities underscore why an assessment of AI maturity is crucial. Rather than ranking operators, a holistic maturity assessment reveals what is possible at each stage of AI maturity, including what kinds of efficiency gains are realistic, how they can be measured credibly and what barriers need to be overcome.

# 3.1 How to measure AI and energy maturity

The GSMA AI Maturity Roadmap provides a structured way to interpret how operators can move from pilots to system-wide impact. It sets out the typical capabilities needed at each level of maturity across four domains: RAN, core/transport, data centres and operations. Rather than ranking operators, the roadmap highlights the practical steps that make AI-driven efficiency measurable, scalable and durable.

Our interviews suggest that most operators are at Level 2 of the GSMA AI Maturity Roadmap, with some approaching Level 3. They have proven that AI works in specific domains, such as energy savings in the RAN or cooling efficiencies in data centres, but they struggle to embed pilots in standard operations. The main barriers to achieving Level 3 seem to be governance gaps and data quality issues. Only a handful of operators are exploring capabilities at Level 4, like energy grid-aware scheduling.

**Table 5** allows operators to identify where they already have capabilities, where gaps remain and what is required to move to the next level of AI maturity, together with their corresponding energy efficiencies.

**Table 5** | GSMA AI Maturity Roadmap: typical capabilities and energy levers by domain and level

| Domain | Level 1 Awareness and foundations | Level 2 Exploration and early adoption | Level 3 Integration and operationalisation | Level 4 Transformation and innovation |
|---|---|---|---|---|
| Typical energy impact | **0-5% (Housekeeping)** | **10-25% (Local optimisation)** | **15-30% (Coordinated operations)** | **15-30% (System-wide)** |
| Primary challenges | **Data quality and consistency** | **Building a business case and securing budget** | **Governance and change management** | **Grid access and market participation** |
| Radio access network (RAN) | **Access network foundations**<br>Site and equipment inventory data is clean, consistent and accurate, and data feeds from networks and energy meters are reliable. | **AI traffic learning**<br>Initial experiments show how AI can identify patterns in traffic and energy data and safely power down radios or carriers when demand is low without affecting the overall service quality. | **Closed-loop orchestration**<br>Traffic learning is now embedded as operational policy to automatically steer traffic, balance service quality and reduce energy across the entire network with minimal human intervention. | **Grid-aware orchestration**<br>AI dynamically orchestrates spectrum and capacity in response to renewable energy availability, grid conditions and storage states. |

Continued on following page >

| Domain | Level 1<br>Awareness and foundations | Level 2<br>Exploration and early adoption | Level 3<br>Integration and operationalisation | Level 4<br>Transformation and innovation |
|---|---|---|---|---|
| Core and transport | **Baseline visibility**<br>Inventory and telemetry data is cleansed and normalised to provide a clear understanding of baseline energy use by component. | **Smart scaling**<br>AI pilots demonstrate savings by automatically switching off idle ports or links during off-peak hours. | **Cross-domain coordination**<br>AI coordinates across core and transport domains to predict traffic patterns and scaling requirements, consistently lowering energy consumption while maintaining service quality. | **Clean power alignment**<br>AI automatically schedules batch processing jobs and maintenance windows to align with renewable energy availability, reducing peak grid demand and marginal carbon emissions. |
| Data centre cooling | **Data centre metering**<br>Separate IT and cooling meters establish clear data flows, allow analytics to run airflow and containment checks to identify obvious waste. | **Single-site optimisation**<br>ML pilots use data to automatically adjust cooling systems for maximum efficiency, demonstrating typical energy savings of 10–15% in a single site. | **Multi-site optimisation**<br>AI is deployed across a portfolio of data centres to predict equipment failures and coordinate cooling strategies across multiple sites to maintain stable, efficient PUE as workloads scale. | **Energy system integration**<br>AI integrates on-site thermal storage and renewable generation with data centre operations, enabling participation in energy flexibility markets where regulations permit. |
| Operations and maintenance | **Basic systems monitoring**<br>Alarms are consolidated, basic energy KPIs are set, and anomalies are flagged for manual “find-and-fix” investigations. | **Basic AIOps**<br>Artificial intelligence for IT operations (AIOps) pilots predict anomalies, future maintenance requirements and optimise support tickets in key functions to demonstrate potential for energy and cost savings by reducing avoidable site visits and equipment failures. | **Governed automation**<br>AI is embedded in automated, policy-driven workflows with verification and rollback mechanisms, ensuring that energy savings and operational improvements are maintained in software and application updates and team changes. | **Circular operations**<br>AI optimises equipment life cycles, extending useful life, enabling refurbishment programmes and minimising operational energy use and e-waste. |

Source: GSMA AI Maturity Roadmap and operator interviews.

## Interpreting the maturity levels

### Level 1: Awareness and foundations

Operators at this stage are focused on getting the basics right – clean data, working meters and simple housekeeping fixes – to build their AI capabilities. For example, in an interview, a European operator described inconsistent asset naming as the main barrier to progress, underlining why data hygiene is a critical first step. In emerging markets, operators emphasised that consolidating asset and energy data across thousands of sites can take up to 18 months due to current levels of fragmentation.

### Level 2: Exploration and early adoption

Here, operators prove that AI is valuable in specific domains where data foundations are sound and energy baselines have been established. For example, one operator in Asia reported that it had achieved approximately 16% reductions in energy use across base stations and data centres using its AI platform, while another European operator saw single-digit savings from AI-assisted RAN optimisation layered on top of the network's existing static control features. These pilots demonstrate that operators are seeing value, but it is largely isolated and at a small scale.

### Level 3: Integration and operationalisation

At this stage, the focus shifts from experiments to embedding AI as policy. Energy-saving features are integrated in access and core networks, as well as operational activities, and safeguarded by appropriate guardrails. One multinational operator noted that many pilots stall at Level 2 without strong governance, while a European solution provider highlighted operational deployments at scale, like predictive cell energy management (PCEM), which delivered approximately 6% power savings across 2,400 sites in Europe. The keys to integration are staff adoption and durability of AI software updates and network infrastructure changes.

### Level 4: Transformation and innovation

At this level, AI is capable of orchestrating internet-connected software and hardware components of a telco network as part of the wider energy system. One operator in Asia described how it is piloting intent-based operations to commercialise the use of natural language to control its networks by 2025. Another European operator is already confronting grid congestion in its key markets, showing why grid-aware scheduling and flexibility participation will matter. A solution provider frames this level as a shift to AI-native, intent-driven programmable networks in which energy efficiency, resilience and flexibility come together across all infrastructure and operations.

# 3.2 Ensuring speed and scale

The pace at which AI can deliver measurable energy savings depends on both the maturity of an operator and the domains in which it is applied. Our interviews show that the strongest early results come from small-scale pilots. However, scaling these gains across networks is slower as issues of data quality, interoperability and organisational trust come into play.

## 3.2.1 Early gains and the steps needed to scale

Operators we interviewed report that early AI pilots show clear promise, demonstrating feasibility and some initial payback, but they lack the scale needed to influence mainstream operations.

Scaling requires more than replicating a pilot in a new system, department or business. For example, a European operator described how a nighttime band switch-off pilot saved energy in parts of its network, but the gains risked being lost as new 5G deployments increased network density and raised overall demand. Another operator stressed that scaling AI pilots depends on initiating new governance frameworks, enabling teams to trust AI-driven changes knowing there are robust guardrails and roll-back mechanisms.

## 3.2.2 What savings are realistic?

The experiences described by interviewees suggest a pattern in the expected range of savings from AI-enabled energy efficiency as their level of maturity increases. These savings represent incremental gains from AI on top of an operator's existing energy baseline, which varies widely between operators and regions.

For instance, even though their AI journey is in its earliest stage, some operators at Level 1 may still have captured significant cost and energy savings from basic measures like equipment upgrades and network modernisation. Others, though, may be starting from a lower baseline with much more room to make similar savings. Conversely, operators at Level 4 may be more likely to have exhausted non-AI efficiency options, meaning their AI gains come on top of an already-efficient foundation. Typical savings over and above the baseline may be:

**Level 1**

**0–5%**
from hygiene and housekeeping actions

**Level 2**

**10–25%**
from savings in the targeted domain, but only…

**5–15%**
network-wide once aggregated

**Level 3**

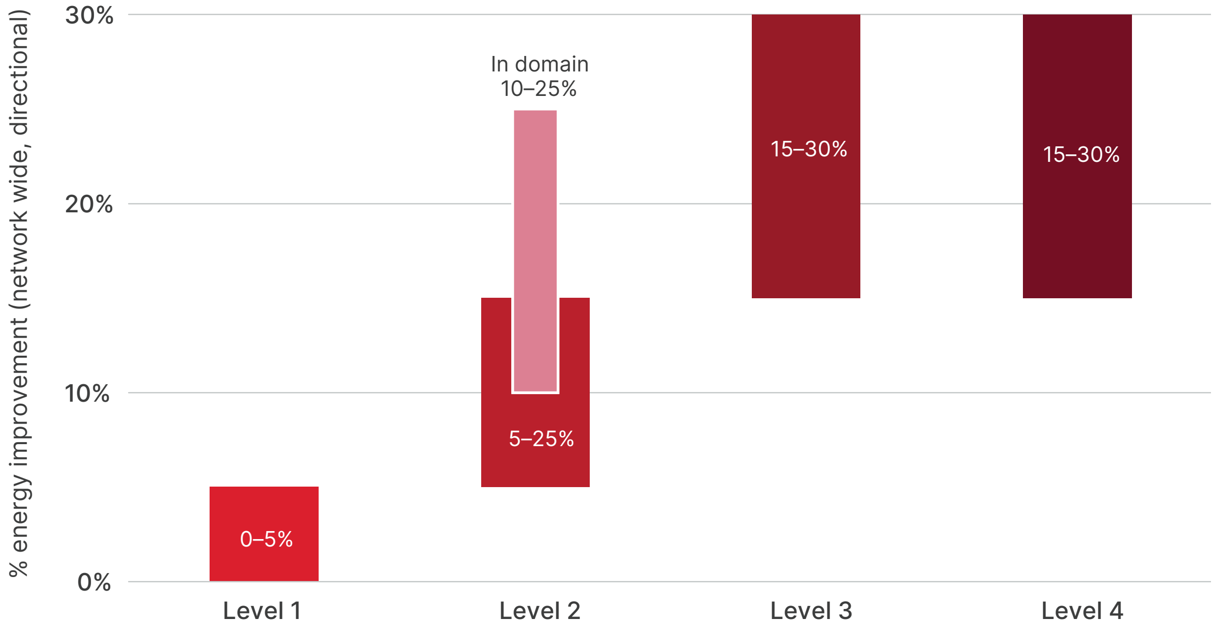
**15–30%**
in sustained reductions as AI policies are embedded

**Level 4**

**15–30%**
with potential for additional reductions and value creation, not only in direct energy savings, but also through grid-aware scheduling, flexibility markets and integration with renewables

Assuming all operators start with a similar baseline, **Figure 5** shows the expected efficiency gains as operators mature.

**Figure 5** | Expected efficiency gains by maturity (directional ranges)




Note that ranges are indicative and not additive across levels.

Source: GSMA Intelligence framing and metered domain results.

## 3.2.3 Case studies

The following case studies illustrate how telco operators claim to be using AI to achieve cost savings and energy efficiencies.

### RAN

- **Vodafone UK with Ericsson:** Employing an AI suite that dynamically puts radios into "deep sleep" reduced daily power use by up to one-third, and by as much as 70% in very low-traffic hours without affecting customer experience.[54]
- **KDDI with Nokia:** Trials of AI-based energy efficiency reduced cell site power by up to 50% at low traffic, averaging around 20% per site.[55]
- **Ericsson global reference:** Using ML to automate sleep modes delivered 14% lower site energy use compared to manual settings, while maintaining quality of service KPIs.[56]
- **Turkcell with P.I. Works:** AI optimisation reduced RAN energy by 3.6% while improving service quality.[57]
- **Indosat Ooredoo Hutchison with Nokia:** A nationwide AI solution that now idles radios automatically at low demand, delivered via the cloud.[58]
- **Bharti Airtel:** Using AI/ML to analyse cell traffic and schedule energy-saving actions.[59]

### Network-wide

- **Comcast:** Since 2019, electricity use per byte has fallen by 49%, and total network energy dropped 11%, achieved through a mix of AI-based operations and virtualisation.[60]

### Data centre cooling

- **Virgin Media O2 with EkkoSense:** Cut cooling energy by 15% across 20 data centres, saving around £1 million per year and avoiding 760 tonnes of $CO_2$.[61]
- **Google DeepMind:** Reduced cooling energy by up to 40% in data centres, improving power usage effectiveness by 15%.[62]
- **Telehouse with EkkoSense:** Achieved a 10% reduction in cooling energy at a data centre in London.[63]

### System-level

- **Elisa:** Repurposed backup batteries at mobile sites into a distributed energy storage system with 150 MWh capacity. These facilities help to cut peak demand at mobile sites and enable them to participate in the grid's flexibility markets, showing how AI can create value beyond direct energy savings.[64]

54 Ericsson (11 March 2025) Vodafone UK and Ericsson trial AI solutions for improved 5G energy efficiency. Press release.
55 Nokia (n.d.) How KDDI used AI to cut RAN energy consumption in half.
56 Ericsson (2019) Automating MIMO energy management with machine learning and AI.
57 GSMA (6 January 2020) Case Study: Turkcell & PI Works.
58 Nokia (7 July 2025) Indosat Ooredoo Hutchison and Nokia partner to reduce energy demand support AI-powered, sustainable operations.
59 GSMA (2024) Bharti Airtel – RAN energy saving with AI and ML.
60 Comcast (7 July 2025) Comcast's Network Traffic Increased More Than 75% While Energy Efficiency Nearly Doubled Over Past Five Years.
61 Virgin Media (18 September 2024) Virgin Media O2 works with EkkoSense to unlock data centre cooling energy savings worth over £1 million per year.
62 Evans, R. and Gao, J. (20 July 2016) DeepMind AI Reduces Google Data Centre Cooling Bill by 40%". Google DeepMind Blog.
63 Billson, C. (10 November 2022) How Telehouse will achieve 461 tonnes reduction in $CO_2$ carbon emissions with EkkoSense. EkkoSense.
64 Elisa (n.d.) Virtual power plant (DES).

# 3.3 Common barriers and practical solutions

While AI has potential to improve energy performance, the operators we interviewed emphasised that progress is slowed by recurring obstacles. Most of these are not technical in nature. Instead, they are organisational and include financial and supply-chain issues that require coordinated solutions across teams, business units and vendors. **Table 6** summarises the most common barriers identified in interviews, ordered by how frequently they were raised, along with the approaches that operators and vendors are using to overcome them.

**Table 6** | Common barriers and how operators are addressing them

| Barrier | Solution (illustrative examples) |
|---|---|
| **Poor data quality:** Inconsistent equipment records, missing energy meters and separate databases prevent AI from accessing data and identifying the repeated patterns needed to make predictions or enhance decision-making. | **Ensure that data is ready for AI:** Reconcile equipment inventories, standardise data formats and install meters where they are missing. Several operators said they have created centralised energy data dashboards and platforms – shared repositories that combine energy, traffic and equipment data to provide a holistic and reliable view across the network. |
| **Difficulties scaling from small pilots to full deployments:** Energy savings proven in small trials do not get embedded in day-to-day operations. | **Strengthen AI-related processes, policies and overall governance:** Ensure that AI initiatives can be embedded in operational workflows with clear approval processes, policies and oversight. For example, one vendor's AI solution has been scaled across thousands of sites and includes built-in performance safeguards.[65] |
| **Lack of trust in autonomous AI decisions:** Teams are reluctant to let AI make changes or take decisions without human oversight or the ability to reverse actions. | **Build an AI-first culture to improve awareness and understanding, and train people in responsible AI use:** One vendor is working with an operator to test an AI "agent" that provides recommendations for optimising capacity and power consumption of the RAN. These outputs can be reviewed by operators before they are implemented, building confidence gradually.[66] |
| **Managing equipment from multiple suppliers:** Networks use hardware and software from different vendors, making it difficult to optimise energy and related services across the entire system. | **Deploy interoperable platforms:** For example, P.I. Works's "Evo platform" – an integrated and open platform for managing mobile networks – is used by multiple operators around the world. [67] |
| **Difficulties in proving actual energy savings:** It can be hard to demonstrate the impact of AI when basic energy-saving features already deliver efficiencies or when it is unclear which improvement caused which saving. | **Track network and cooling energy separately:** Only credit AI when it replaces a manual process or a function executed on a fixed time schedule. Several operators now require measurement of kWh saved while keeping service quality metrics constant. |
| **Budget constraints and old equipment:** Ageing infrastructure and limited capital budgets slow adoption of new AI capabilities. | **Prioritise software upgrades for existing hardware:** Many AI-enabled energy features (such as intelligent sleep modes and workload prediction) can be activated through software updates without replacing physical equipment. |
| **Power grid constraints in some regions:** In certain markets, electricity availability and location restrictions limit where new infrastructure can be built. | **Extend AI coordination beyond the network itself:** Use AI to schedule high-energy tasks when renewable power is available, manage battery storage and participate in programmes that reward flexible energy use. For example, using AI to coordinate virtual power plants is an approach that could be emulated by telcos.[68] |

65 Ericsson (12 May 2023) Ericsson success with FutureNet World Network Sustainability Award.
66 Ericsson (5 March 2025) Telenor and Ericsson successfully co-create agentic AI for radio controls and expand innovation partnership.
67 See: https://piworks.net/about/customers. Other similar platforms and services are available from different vendors.
68 Elisa Virtual Power Plant (DES).

These barriers explain why it is difficult for operators to advance from Level 2 to Level 3 of the maturity roadmap and beyond. Solutions are emerging, from centralised data platforms and more consistent metering to explainable AI tools and improved interoperability across vendors and systems. However, implementing these solutions at scale requires changes in governance, processes and capability, as well as technology.

Many operators we spoke with emphasised that progress depends on building new skills in data management and engineering, AI operations and cross-functional collaboration between network, IT and energy teams. As one operator put it: “AI doesn’t fail in the lab; it fails in the change-control process.”

# 3.4 Governance and guardrails

The barriers described in Section 3.3 point to a common theme affecting all enterprises seeking to implement AI: operators struggle less with proving that AI works and more with making it work reliably, safely and at scale in live networks and complex operations. This is where governance becomes essential. In our interviews, several operators described promising AI pilots that had stalled because operational teams lacked the confidence to deploy solutions without clear policies and safeguards.

## 3.4.1 The trust gap

Network operations teams are accountable for service quality and uptime, so they need to be confident that AI-driven energy optimisation and similar solutions will not diminish the customer experience or cause unexpected failures in the network. Our interviews revealed three types of safeguards essential for building this confidence:

**1. Transparency and explainability**

Operators need to understand why AI recommends specific actions and whether those actions align with both energy and service quality goals. One multinational operator highlighted the value of AI systems that provide clear reasoning for their recommendations, allowing them to validate decisions before implementation. One vendor described its collaboration with an operator, using a framework that allows humans to review and, if necessary, override AI decisions, embedding accountability and human experience while still capturing efficiency gains. Another operator provides users of its internal chatbot capability with immediate visibility of prompting and output costs, encouraging users to make appropriate AI model selections, which could save energy as well as money.

**2. Verification and rollback**

Trust also requires being able to demonstrate that energy-saving actions will not diminish the customer experience or compromise network performance. Several operators stressed the importance of automated safeguards that monitor service quality metrics continuously. If performance thresholds are threatened, the system automatically reverts to the previous configuration. For example, Ericsson's PCEM incorporates this approach, enabling energy features to be deployed across thousands of sites while maintaining service reliability.[69] This "fail-safe" design gives operational teams confidence to scale AI beyond small pilots.

**3. Consistent measurement and reporting**

Without credible traffic and energy baselines, claimed savings risk scepticism from both internal and external stakeholders. In interviews, operators argued for tracking energy consumption by domain (e.g. RAN vs cooling) and reporting energy savings alongside quality-of-service metrics. This dual measurement approach could prevent efficiency being achieved at the expense of user experience and avoid double-counting when multiple solutions operate simultaneously.

69 Ericsson (12 May 2023) Ericsson success with FutureNet World Network Sustainability Award.

## 3.4.2. Industry-wide standards

Beyond individual practices, several operators suggested there is a role for industry-wide coordination. The GSMA Responsible AI Maturity Roadmap[70] provides a foundation for this coordination, establishing principles of fairness, human agency and oversight, privacy and security, safety and robustness, transparency, accountability and environmental impact. Building on this framework, the GSMA could help by:

- **Convening vendors and operators** to align on interoperability standards, enabling AI optimisation to function across equipment from multiple suppliers.
- **Establishing standard definitions** for what qualifies as AI-driven energy savings versus conventional efficiency measures.
- **Developing voluntary reporting frameworks** that present energy reductions and service quality indicators together, building credibility with regulators and stakeholders and aligned with responsible AI principles.

Although often viewed as additional bureaucracy, governance mechanisms like transparency, verification and credible measurement are an essential foundation that will help operators move from small-scale experiments to system-wide AI implementation. Without them, operational teams are likely to remain rightfully cautious, and the promise of AI technologies, although demonstrated in the pilot phase, remains locked up. Several operators emphasised that building governance frameworks early, even before AI capabilities are fully mature, accelerates deployment later by establishing principles of trust and clear accountability.

The technical capability to optimise energy use with AI already exists in many cases. What determines whether operators can capture those benefits at scale is whether, alongside the technology, they have built the organisational infrastructure and culture – the policies, processes and safeguards – to deploy AI confidently across their networks. This is the key difference between operators at Level 2 working on isolated pilots and those progressing to Level 3 and beyond, where AI-driven efficiency becomes standard practice.

70 GSMA Responsible AI Maturity Roadmap: A Step-by-Step Guide

# 04 Future trends and strategic implications

AI is changing energy use in telecommunications, and this report has examined the opportunities, barriers and early lessons from operators deploying AI-driven solutions. This section looks to the future to consider how the energy footprint of AI might evolve, which operators face the greatest exposure and what actions can position the telco sector to manage uncertainty over the next five years.

## 4.1 AI is disrupting the notion that data traffic is predictable

Earlier sections of this report showed how AI creates both opportunities and new energy demands. What is changing is both the pattern of that demand and its magnitude. Until now, telcos have planned their energy needs around predictable network behaviour, assuming steady year-on-year traffic growth, gradual infrastructure upgrades and reliable efficiency improvements. AI is disrupting this model by making traffic patterns and network demand harder to predict. For instance, through:

- **Demand that moves around:** AI workloads tend to shift between data centres, networks and end user devices depending on where models are trained and where they run. Energy-intensive model training is typically concentrated in hyperscale data centres, but inference can occur anywhere, from the same hyperscale clouds to edge data centres and even smartphones. This distribution makes it harder to predict where electricity demand will emerge and when.

- **Traffic that behaves differently:** While AI-related data still represents a small fraction of total mobile traffic, early analysis suggests it has distinctive characteristics. It is typically uplink-heavy, requires low-latency connections for its applications and typically occurs in bursts rather than being uniformly distributed. These new patterns limit opportunities for traditional energy-saving techniques like cell-site sleep modes, which work best when traffic is more predictable and off-peak periods are quieter.

- **Hardware gains that do not keep pace:** Individual computing components are becoming more efficient. For example, GPUs, task-specific chips and even entire cooling systems now deliver more performance per watt than previous generations. However, energy-per-bit calculations mask the total growth in computing workloads, which could, in some circumstances, outpace these improvements. This is the "rebound effect", which has been observed in similar industrial and digital transitions. One operator noted this dynamic, observing that network modernisation had reduced per-bit energy usage, but total consumption continued to climb as data volumes across their network surged.

- **Variability that increases:** The combination of rising compute workloads, distributed inference and fluctuations in the supply of renewable energy means that operators must manage variability as well as volume. Energy profiles are therefore becoming less predictable, keeping operating costs high and complicating network capacity planning and energy grid coordination. Two operators flagged this challenge, noting that grid constraints in some markets now limit where they can expand infrastructure, forcing them to think about energy availability before site selection.

# 4.2 How do these changes affect operators?

Not every operator will experience the energy impact of AI in the same way. The level of exposure depends on their maturity level, infrastructure ownership and mix, business model and local energy grid conditions. Nevertheless, three broad categories emerged from our interviews:

- **AI-anchor operators:** These operators are investing directly in AI infrastructure, either through operator-owned data centres or enterprise colocation facilities that host third-party compute. They face the most immediate risks associated with growth in electricity demand as AI workloads expand, such as direct power costs, cooling infrastructure requirements and potential grid capacity constraints. Examples include telcos in Asia and the Middle East, which are making data centre investments that specifically target AI workloads.[71] To mitigate risks, these operators can co-locate facilities in regions with more reliable renewable energy supplies, negotiate long-term PPAs or use innovative cooling methods to handle new high-density GPU racks in their data centres.[72]

- **AI-carrier operators:** These operators rely primarily on hyperscalers for their compute requirements but carry the traffic that connects users to these cloud-based AI services. They feel pressure mainly through radio access and transport layers as uplink demand increases. Examples include most European and North American mobile operators, alongside many operators in Latin America and Asia.[73] They host minimal AI compute themselves, primarily for network optimisation and improved customer service operations, but see traffic patterns changing as users adopt AI-powered applications from Big Tech. To mitigate the risks, these operators can explore more sophisticated traffic prediction algorithms to anticipate burst patterns and deploy AI-driven network optimisation to offset overall growth in network load.

- **AI-emerging operators:** These operators have limited AI footprints, which means minimal data centre investment and participation in markets where AI applications are only slowly being adopted by consumers and businesses. They face lower near-term energy risks but could be caught unprepared if AI features become more common across devices and applications. To mitigate the risks, operators can ensure they build foundational capabilities now, such as data quality, metering infrastructure and basic AI operations skills, before AI traffic becomes a material proportion of overall traffic. Several operators we interviewed considered themselves in this category and described investments in Level 1 and Level 2 maturity as much-needed insurance against future uncertainty.

The differences between these categories are subtle, although some differences are already visible in energy trends. AI-anchor operators with significant data centre operations report that cooling and IT loads in their facilities now represent 35–40% of total electricity consumption, a significant increase from just five years ago.[74] AI-carrier operators see modest increases in RAN energy consumption due to increased 5G network density and mobile data traffic continuing its exponential growth. AI-emerging operators often show declining total energy as 2G and 3G shutdowns, fleet electrification, cable replacement works and hardware refreshes offset modest growth from traffic.

As illustrated in **Figure 1**, by 2028, these patterns are expected to diverge more sharply. AI-anchor operators could see data centre electricity demand triple from 2024 levels. AI-carrier operators might experience lower overall energy costs than AI anchor operators, but a more marked rise in network energy without more effective and coordinated AI-driven optimisation. AI-emerging operators face the risk of rapid catch-up if AI adoption accelerates in their markets – a scenario that leaves little time to develop wider organisational preparedness.

71 Ericsson (12 May 2023) Ericsson success with FutureNet World Network Sustainability Award.
72 GSMA Responsible AI Maturity Roadmap
73 For example, see: Total Telecom (7 August 2025) Saudi's center3 unveils plans for 1 Gigawatt of data centre capacity by 2030 and Gao, J. and Liang, J. (1 April 2024) Major Chinese telecom companies slow infrastructure investment, target AI computing power. Digitimes Asia.
74 AtlasEdge (n.d.) AtlasEdge Wins the Prestigious Green Apple Environment Award!
75 For example, several European operators are suggesting that Big Tech should shoulder the required investments in connectivity arising from demand for AI. See: Wooden, A. (14 February 2022) European telcos say big tech should help fund network infrastructure (again). Telecoms.com.

# 4.3 What does the future hold?

Predicting how AI will shape energy use by telcos in 2030 is difficult. Adoption rates, technological evolution, regulatory frameworks and customer preferences are all in flux. Rather than attempting to make a precise forecast, operators can understand their potential exposure by examining how several key drivers might combine to create different futures, including:

- **Pace of AI adoption:** How quickly AI is adopted and becomes embedded in consumer applications, enterprise services and network operations has already proved unpredictable. For example, ChatGPT reached 100 million users in two months and now sees more than 5 billion visits monthly, defying every precedent.[75] Whether this pace continues, accelerates or plateaus will alter the energy trajectories for operators.

- **Inference location:** Where AI processing occurs matters as much as volume. While model training usually takes place in large hyperscaler facilities, inference can occur anywhere, from cloud data centres to edge locations, and even smartphones. Traffic patterns are more uplink-heavy when AI processing is cloud-based but may become more distributed as edge and on-device capabilities mature. Each pattern creates different energy profiles and accompanying planning challenges for operators.

- **Organisational readiness:** Technical solutions tend to deliver value only when operators build in the governance, skills and coordination mechanisms to deploy them effectively. The gap between proving AI works in pilot projects and embedding it in operations can take two to three years (or more) to bridge, often longer than the technical implementation itself. Operators that start building these capabilities early can gain years of advantage over those who wait.

- **Grid and regulatory context:** Energy availability increasingly constrains network and data centre infrastructure decisions. Two operators in our interviews noted that grid capacity now determines where they can expand sites, forcing energy planning ahead of site location planning. Renewable energy integration, dynamic pricing and regulatory pressure on emissions all influence which efficiency strategies become viable and when.

These drivers are likely to interact in complex ways. For example, rapid AI adoption without organisational readiness may create growth in energy that overwhelms efficiency gains. Or investing in stronger grid coordination capabilities may return limited value if the increase in AI-related traffic remains modest. While the interplay between them could produce myriad scenarios, it is helpful to imagine a small number of more plausible ones based on our interviews and current trajectories.

76 Goldman Sachs (4 February 2025) AI to drive 165% increase in data center power demand by 2030.
77 FirstPageSage (3 October 2025) ChatGPT Usage Statistics: October 2025.

## 4.3.1 Scenario 1: Steady evolution

In this scenario, AI adoption is gradual until 2030. Most operators see AI traffic reach 15% or 20% of total mobile data, concentrated in specific applications rather than embedded everywhere. Data centre electricity demand roughly doubles from 2025 levels, but most growth is experienced by hyperscalers. Consequently, network energy consumption continues to remain roughly flat as efficiency gains from ongoing 2G/3G shutdowns and network modernisation largely offset new loads.

In this scenario, operators have time to build capability in a deliberate fashion. Measurement infrastructure, data platforms and pilot AI tools progress without urgent pressure. The challenge shifts to maintaining momentum when business cases for AI and upgrades remain marginal and competing priorities dominate attention. Several operators we interviewed considered this their working assumption, describing energy management as important but not yet critical.

| How do different operators fare? | | |
|---|---|---|
| **AI-anchor operators** | **AI-carrier operators** | **AI-emerging operators** |
| likely to see returns on data centre investments materialise only slowly, with infrastructure running below expected usage levels. | likely to benefit most, as their lighter capital commitments and complementary relationship with hyperscalers aligns well with the gradual growth in AI adoption, and traffic patterns remain manageable through conventional optimisation techniques. | likely to be able to justify their more cautious stance, although foundational capabilities like data quality, cross-domain coordination and basic AI operations skills still need to be built if demand accelerates. Those who treat this five-year period as preparation rather than pause are likely to position themselves for faster transitions if adoption quickens. |

## 4.3.2. Scenario 2: Mainstream transformation

In this scenario, AI becomes completely embedded in core telco processes and consumer services. Traffic with AI characteristics (uplink heavy, burst oriented, low latency) reaches 20–50% of total mobile data. However, data centre electricity demand grows by three to four times current levels as operators expand their edge infrastructure and enterprise services. Without a coordinated response, network energy could rise by 5–15%.

This is the scenario where AI-driven optimisation becomes essential rather than experimental. Operators need to reach at least Level 2 or 3 maturity by 2027 to manage the unpredictable nature of AI data traffic. This requires governance frameworks to support the use of embedded AI features to change legacy systems and approaches. Additional verification systems are also needed, some with humans in the loop, to monitor service quality alongside network efficiency. To do this effectively, operators will need cross-functional teams with both network operations and data science expertise.

In this future, wider organisational capability becomes the binding constraint. Technical solutions exist but deploying them at scale takes longer than expected. Operators that started building governance structures, training programmes and vendor partnerships two years earlier reach deployment faster than those responding reactively. The winners in this scenario, who are able to use AI effectively to optimise energy consumption, are those that bridged the gap between pilot and production most effectively.

**How do different operators fare?**

| AI-anchor operators | AI-carrier operators | AI-emerging operators |
|---|---|---|
| see strong demand growth that validates their original infrastructure investments, although cooling and power costs climb faster than anticipated, requiring them to adopt more sophisticated energy procurement strategies and grid partnerships. | face the sharpest transition pressure as uplink growth and burst traffic patterns stress current RAN infrastructure, while legacy business models remain focused on connectivity rather than compute. Even though network energy usage remains largely unaffected, these operators need sophisticated AI-driven network optimisation at scale to maintain margins. | experience some catch-up pressure if AI adoption accelerates in their markets, leaving them with limited time to develop their organisational preparedness. Those that invested in Level 1 foundations early can adapt faster than those starting from scratch. |

Overall, interactions with the energy grid must become more sophisticated, in the same way that energy system operators must collaborate more effectively with telecommunications and other infrastructure providers. All types of telecommunications operators will need to align network planning with energy availability forecasts, negotiate more flexible PPAs and begin exploring new ways to respond to increasing traffic demand pressure. For everyone, energy shifts from being a cost that is managed to a strategic factor influencing site selection, network and IT architecture choices and service offerings.

### 4.3.3 Scenario 3: Pervasive disruption

In this scenario, the much-hyped agentic AI capabilities and near-real time inference workloads become common. AI traffic exceeds 50% of total data, with continuous interaction patterns that eliminate traditional quiet periods when cell-site sleep modes used to work. Data centre power consumption grows more than fourfold from current levels as edge computing proliferates, and network capacity must grow accordingly. Without full orchestration and grid coordination, baseline network energy could rise 15–25%.

In this scenario, Level 4 maturity is necessary for operational viability. Operators implement grid-aware scheduling that shifts compute workloads based on the availability of renewable energy and its dynamic pricing. Energy management becomes fully integrated with network control, and systems optimise continuously across traffic demand, service quality, energy cost and carbon intensity to ensure financial and sustainability targets are met.

The organisational requirements are substantial. For example, teams need skills in AI operations, energy markets and grid coordination. Vendor partnerships must extend beyond traditional equipment supply to include energy analytics and coordination platforms, and common open standards for cross-vendor orchestration are no longer optional.

**How do different operators fare?**

| AI-anchor operators | AI-carrier operators | AI-emerging operators |
|---|---|---|
| become energy-intensive businesses requiring fundamental repositioning. This is because data centre electricity now represents 25% or more of total consumption, demanding sophisticated cooling solutions, long-term renewable PPAs and participation in grid flexibility markets. Those co-located in areas with reliable renewable or non-fossil-fuel energy supplies gain competitive advantage. | must deploy full AI orchestration and traffic prediction to manage continuously varying loads on the network, with energy cost becoming a primary operational metric alongside quality of service. | face existential pressure if they have not been able to build foundational capabilities. Catching up becomes nearly impossible as the gap between pilot and production capability at the extreme loads experienced requires years, not months, to bridge. |

This scenario also creates new opportunities for some but not all operators. For example, those with excess network capacity can participate in demand response markets, generating revenue from grid flexibility. Renewable integration becomes a competitive advantage in markets with strong decarbonisation commitments. In these ways, energy expertise shifts from being an expense to a strategic differentiator.

### 4.3.4 What do these scenarios mean for planning?

These scenarios should be viewed as planning tools rather than predictions. The future will likely combine elements of each, with variations by market region and operator type. Their value, however, lies in revealing common themes and key differences that can support planning.

Across all three scenarios, for example, reliable measurement infrastructure and organisational capability create long-term value. Whether AI adoption remains modest or surges ahead, operators without clean data, governance frameworks and skilled teams will struggle to respond effectively. These foundations can take years to build, making them priorities regardless of which scenario unfolds.

However, the timing of investments varies dramatically between scenarios. In the steady evolution of AI, operators can phase capabilities over five to seven years with little pressure to change faster. In mainstream transformation, however, that window compresses to three years. In the pervasive disruption of AI, operators need Level 3 to Level 4 capabilities operational by 2028–2029, which means they must start preparing now.

In practice, the scenarios show that adaptability matters more than prediction. Operators that manage uncertainty most effectively will be those that review future scenarios regularly and update their planning assumptions as new data emerges. This will help them identify which investments are most likely to create value in multiple futures rather than trapping them in one trajectory. They will view AI and energy management as a process requiring continuous evolution, not a problem to solve just once.

# 4.4 What operators should do now

The operators best positioned to manage the energy impact of AI will combine governance, culture, skills and industry partnerships with key technical capabilities. Based on our interviews and analysis, four priorities stand out, each grounded in the GSMA AI Maturity Framework and lessons highlighted in Sections 2 and 3.

## 4.4.1 Build reliable data and improve observability

Like many successful enterprises in other sectors, operators we interviewed that have progressed beyond Level 1 emphasised that reliable data is the prerequisite for everything else. Without clean baselines, AI optimisation cannot demonstrate energy benefits, pilots will remain isolated experiments and investment cases will just be speculative.

**Recommendations for operators**

- **Separate energy accounting by domain:** Track network, data centre and auxiliary systems independently. Several operators noted that this distinction helps pinpoint energy-intensive areas and enables targeted interventions that reduce overall energy consumption.
- **Instrument systematically:** Deploy energy meters where they are missing, particularly in older RAN sites and colocation facilities. As one operator said, inconsistent metering was their single biggest barrier to AI deployment; they could not optimise energy use that they could not measure.
- **Create shared data platforms:** Build repositories that combine energy, traffic, service quality and asset data in formats that both network operations and sustainability teams can use. Multiple operators described using enterprise data platforms, such as data lakes, as the foundation for moving from Level 1 to Level 2 and beyond, giving AI solutions consistent inputs to drive energy efficiency across the network.
- **Report metrics together:** Present energy savings and quality of service indicators together, establishing from the outset that efficiency gains cannot come at the expense of customer experience. This dual accountability builds credibility and strengthens commitments to sustainable operations.

## 4.4.2 Move from pilots to production

The gap between Level 2 and Level 3 maturity is where operators say they are likely to stall. They have proven that AI works in certain domains but cannot scale up those successes. Despite numerous opportunities from advancements in AI capabilities, this is mainly an organisational and cultural barrier.

**Recommendations for operators**

- **Formalise governance:** Create change-control pathways specifically for AI-driven energy optimisation features, with clear approval processes, roll-back mechanisms and lines of accountability. Recent cases using explainable AI demonstrate that human operators can review recommendations, approve changes and override decisions if service quality metrics are threatened.
- **Build verification frameworks:** Implement automated monitoring that tracks both energy savings and service KPIs continuously. If thresholds are breached, systems should revert to their previous configurations automatically. This fail-safe approach gives operational teams the confidence to deploy AI broadly rather than keeping it in sandboxes.
- **Assign cross-functional ownership:** Operators that scale successfully create permanent teams combining network engineering, data science and sustainability expertise. These teams have dedicated resources, career paths and clear KPIs, which include energy reduction targets.
- **Document and standardise:** Capture what works in formal procedures so that deployments can expand and accelerate progress without requiring the same small expert team to configure every site. For example, scaling can stall without reusable “playbooks” that field engineers can follow to replicate energy-saving configurations.

## 4.4.3 Design for multi-vendor coordination

Real networks inevitably use equipment from multiple suppliers. AI optimisation that only works within single-vendor domains will capture limited energy benefits. Operators that want to reach Level 3 or Level 4 need coordination capabilities that span their entire network and operational footprint.

**Recommendations for operators**

- **Demand open interfaces from vendors:** Push suppliers to support APIs and data standards that enable cross-vendor orchestration. Vendor cooperation can improve when framed as a competitive advantage and an imperative for network-wide energy efficiency.
- **Reduce system complexity:** Use standardised integration techniques and open platforms, and ensure data interoperability across systems, by working with vendors that help with orchestration across multiple network systems to optimise energy use holistically and reduce overall complexity.
- **Start with domains where coordination matters most:** Traffic steering and load balancing, coordination between RAN and core network elements and alignment of cooling systems with growing IT workloads all require different versions of cross-vendor visibility, focusing governance and targeting investment to minimise energy waste.
- **Participate in standards development:** Engage with the GSMA and other industry bodies that are developing interoperability frameworks and maturity assessments. Several operators noted that waiting for the perfect standards to materialise is a mistake, while getting involved early helps shape outcomes and build internal expertise for energy-focused AI deployments.

### 4.4.4 Build skills and capacity

It is well established that technology capabilities change faster than organisational capabilities. Operators that invest only in tools without developing internal expertise are more likely to find themselves dependent on vendors, unable to customise solutions to their specific contexts and slower to adapt as AI evolves.

**Recommendations for operators**

- **Develop internal AI operations expertise:** Train network engineers in basic data science concepts and data scientists in telco operations, with a focus on energy optimisation use cases. The most effective teams will have staff who can bridge both domains.
- **Create continuous learning structures:** Establish regular knowledge-sharing forums, bring in external expertise periodically and allocate time for teams to experiment with new approaches that demonstrate improved energy performance.
- **Link to career development:** Sustainability, AI and energy management skills are increasingly valuable. Make it clear that developing these capabilities creates career opportunities, preventing the risk that expertise remains concentrated in just a few individuals who can become bottlenecks or retention risks.
- **Build vendor partnerships strategically:** Work closely with suppliers on early deployments but insist on knowledge transfer so that internal teams can operate and optimise energy-saving solutions without permanent dependence on vendors.

## 4.5 How to adapt to an uncertain future

AI is reshaping how operators use and manage energy, but its full impact has yet to be felt.

This report began by asking whether AI would help networks reduce their energy consumption or add to it. The answer from operators, vendors and analysis is both, depending on how it is deployed.

AI introduces new electricity demand, particularly in data centres and through variable traffic patterns that limit the usefulness of traditional energy-saving techniques. Yet, it also enables smarter network control, predictive maintenance and system-wide optimisation that would all be impossible with static rules or conventional non-AI approaches.

The telco sector has navigated major technology transitions before, from circuit switching to packet networks and from dedicated hardware to virtualised functions. However, AI creates distinct challenges:

- **The energy impact is bidirectional:** Unlike previous transitions, where efficiency came mainly from hardware replacement, AI simultaneously creates demand through compute-intensive workloads and reduces it through network optimisation. Operators cannot simply upgrade to the latest technology without considering how they will manage this balance.

- **Adoption is happening at a rapid pace:** AI is being deployed faster than other technologies, providing less time for infrastructure planning and capability development. The gap between early pilots and mainstream deployment could be two to three years rather than the five to 10 years typical of past transitions.

- **The stakes extend beyond operations:** AI-driven energy decisions that impact telco networks now intersect with the reliability of the energy grid, the integration of renewables and regulatory frameworks in ways that put operators at the centre of broader energy system challenges.

Whether AI becomes a burden or a catalyst for efficiency depends on the choices that operators, vendors, consumers and policymakers make over the next couple of years, not in the next decade. The sector needs new capabilities, going beyond optimising for efficiency to orchestrating traffic, costs and energy dynamically across networks, data centres and interactions with the wider electricity system. This is Level 4 transformation, which is perhaps becoming necessary faster than many operators might have expected.

The operators most likely to succeed are those that treat AI energy management as a core operational requirement integrated in network planning, investment decisions and sustainability commitments. They recognise that the value of AI comes from augmenting rather than replacing human judgement, ensuring that it provides greater visibility, increased predictive power and wider coordination that manual operations cannot achieve at scale.

Our interviews with operators and vendors suggest that the technical foundations are already being put in place. What remains is building the organisational infrastructure and adaptability, including governance, skills, standards and collaboration mechanisms, and using these effectively alongside the technology.

The window for building these capabilities is narrowing. Operators that start now will be positioned to scale when demand accelerates. Those that wait risk finding themselves without the organisational foundations to deploy solutions quickly. The energy impact of AI on telecommunications will stretch beyond 2030. Successful organisations will have the ability to sense changes, respond quickly and learn from every new shift.

The next few years will determine whether the telco sector shapes the energy trajectory of AI or merely responds to it. Shaping it may require operators to consider bigger questions: do they become active players in the AI value chain, enabling the widespread use of AI or simply providing connectivity for others? The answer may lie in their familiarity with AI, their use of the technology and the pace at which they act.

**GSMA Head Office**
1 Angel Lane
London
EC4R 3AB
United Kingdom