
Telco AI:

State of the Market, Q2 2024

AI in telecoms – in theory and in action

GSMA™

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

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

Follow the GSMA on X: [@GSMA](https://twitter.com/GSMA)

Authors

Peter Jarich
Head of GSMA
Intelligence

Tim Hatt
Head of Research
and Consulting

Sayali Borole
Senior Analyst

GSMA Intelligence

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

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

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

www.gsmaintelligence.com

info@gsmaintelligence.com

Charting the rise of telco AI



Peter Jarich
Head of GSMA Intelligence

I recently took a friend from the US on a tour of some vineyards in France. As she took photos on her phone, AI tools suggested a location – Napa Valley. Why it was so wrong is unclear – her VPN and IP address? Or AI confusing Bordeaux for Napa?

What is clear is the important lessons to be learnt on AI:

- Despite overwhelming amounts of hype, AI – and generative AI (genAI) in particular – is still a maturing technology.
- It is hard to deny AI will play an increasingly greater role in our everyday lives.
- Ensuring AI lives up to its potential will require robust data that is accurate, complete, secure and used in a way that is respectful of personal sensitivities. It will also require recognition that genAI is only a small part of the broader AI landscape.
- As connectivity providers for our smartphones, tablets, homes and connected lives, telcos play a role in how AI evolves and proliferates – both in terms of their own operations and the AI-enabled services they offer.

Against this backdrop, GSMA Intelligence is launching this research series to investigate how telcos are putting AI to use, how they are innovating and what they (and supporting ecosystem players) need to be thinking about as they continue their AI journeys. The series will look at strategy and technology planning, AI sustainability, and the results of a benchmarking exercise establishing how advanced telcos are in their adoption and use of AI.

While rapidly evolving AI use cases have injected energy into the market, AI is no different from any other technology in one fundamental way. No single stakeholder or industry observer has all the answers or has a preternatural understanding of the market's trajectory. This series will form part of the continued dialogue helping the industry to take advantage of AI's potential and to use it responsibly.

About the series

AI is a transformative technology, but it is important to look beyond the hype. This is the first report in a four-part series on AI strategy in telecoms. The research identifies the parts of the innovation cycle that matter most, and how they translate into commercial activity and possible changes in corporate strategy.

Telco AI: State of the Market quarterly series: research approach		
 <p>Market context</p> <ul style="list-style-type: none"> • A recap of critical industry developments and implications • Putting industry progress in context 	 <p>Leader profiles</p> <ul style="list-style-type: none"> • Case studies of telco AI in action • Understanding best practice for the sector 	 <p>Deep dives</p> <ul style="list-style-type: none"> • Deep dives into key topics <ul style="list-style-type: none"> • AI data and systems • energy efficiency • strategy and culture • AI network strategy

In tandem with the report series, GSMA Intelligence is developing an AI benchmark to track AI use and implementation in the telecoms sector. We invite all interested operators and stakeholders across the ecosystem to contact us to participate.

This edition explores the expanding use of AI in the telecoms sector, examining both core AI and genAI. It highlights key use cases, contrasts AI with genAI adoption, and delves into the increasing role of large language models (LLMs) in driving industry advancements.

01

Executive summary

Page 5

02

Market context

Page 8

03

Leader profiles: telco AI in action

Page 17

04

Deep dive: core AI versus genAI

Page 22

05

Deep dive: open versus
proprietary LLMs

Page 28

01

Executive
summary

State of the market

01

The AI potential writ large

AI encompasses a series of overlapping technologies, each suited to different use cases. While genAI has recently grabbed the most attention, core AI has long been instrumental in telecoms. Telcos have been fast adopters of the technology (in contrast to delays embracing previous technology shifts). Their deployments balance pragmatism and ambition. Early wins have concentrated on network operations, with efforts expanding into the services domain.

02

Core AI versus genAI

Core AI has enhanced telecoms operations, driving efficiencies in network management and customer support through tasks such as fault detection, security and traffic management. The technology is now expanding beyond specialised teams to broader organisational use. By contrast, GenAI is at an earlier stage but is progressing quickly through the industry. Only 20% of operators surveyed had deployed it in mid-2023, but this is now likely to be above 90%.

03

Long-term co-existence

For all the hype of ChatGPT and its genAI peers, the reality is that core AI is likely to exist in parallel for many years. This partly comes down to cost, return on investment (RoI) and the complexity of using genAI at scale. It also comes down to the fact that core AI is sufficient for many of the tasks a typical telco would target. GSMA Intelligence research and industry conversations outline how these use cases could play out in practice.

04

Understanding the LLM choice

LLMs are a component of genAI. They focus on text generation, while genAI encompasses a broader range of content creation capabilities. LLMs are revolutionising telecoms by enhancing data processing, customer interaction and network management. To develop scalable telco-specific LLMs, collaborative efforts are underway, including partnerships between Amdocs, Nvidia and Microsoft Azure. The Global Telco AI Alliance is also promoting these advances. However, creating universally applicable LLMs remains challenging, requiring significant refinement and adaptation for diverse telecoms environments.

05

Responsible AI by design, from the start

Data privacy and regulatory compliance are critical aspects of using AI. Operators worldwide are committed to implementing AI responsibly, recognising its risks. Examples include the GSMA's AI for Impact initiative, involving 27 major mobile operators; True Corporation's guidelines for ethical AI usage; and BT's values-led AI vision. Regulations are likely to evolve at a slower pace than technology, which puts the onus on telcos (as with other private sector participants) to have transparent and accountable commitments to responsible AI.

AI in numbers

81%

Percentage of operators globally testing genAI solutions

Commercial deployments at scale are the preserve of a smaller cohort, but most operators will have deployed genAI in some shape or form in the next 12 months – in areas such as network operations support and customer contact centres.

42%

Percentage of operators that see today's early stage of genAI technology development as the top obstacle to deployment (followed by uncertain RoI)

If tech maturity and RoI dominate operator thinking as they plan genAI rollouts, deployments that prove the technology's capabilities and business value will help justify investment.

4%

Potential opex savings if energy costs are reduced by 20%

Energy accounts for 45% of network opex and 15–20% of total operational spend for an average telco. AI is increasingly being used to effect change in RAN sleep states, site location and cooling systems. All can help lower energy consumption.

77%

The response rate for SK Telecom's CS Agent using a telco-specific LLM, compared to 44% with a general LLM (as highlighted at the DTW24 event)

Telco-specific LLMs enable agents to access information from operational support and business support systems, resulting in quicker and more accurate responses to customer queries.

01 Executive summary
Page 5

02 Market context
Page 8

03 Leader profiles: telco AI in action
Page 17

04 Deep dive: core AI versus genAI
Page 22

05 Deep dive: open versus
proprietary LLMs
Page 28

02 Market context

Forces influencing the development and use of AI

AI is a pervasive technology; it has a bearing on many parts of the telecoms industry and other sectors. However, it is easy to get lost in the noise and fail to understand its implications. GSMA Intelligence takes a structured approach to consider the factors influencing AI at the technological, commercial and regulatory levels.

Technological

- **Definitions matter** – The distinction between core AI, machine learning and genAI (terms often incorrectly used interchangeably) matters in terms of level setting how a telco uses AI in its business. A basic delineation of each AI variant as it grows exponentially is key.
- **Build or buy?** The question of whether to source off-the-shelf LLMs or develop in-house is a strategic one depending on use case.

Commercial

- **A transcendent topic** – AI is one of the most important technological shifts since the development of the internet. Understanding the benefits and risks at the sector and company levels is critical to successful implementation. Moreover, it demands transformational changes across company structures (in technology, strategy, culture and talent).
- **Easy wins versus slow burns** – Some AI uses are obvious (e.g. automation), while others are not, or take time to assess (e.g. customer engagement). Measuring AI success is challenging due to difficulties quantifying the benefits across various dimensions.

Regulatory

- **Ethics at the core** – Technology tends to move more quickly than regulation. However, the ramifications of AI mean responsible stewardship must be a priority.
- **It pays to understand the landscape** – The US, EU and UK have made AI a political priority, to get ahead of the curve. Principles have been established; these likely presage specific compliance requirements, and penalties for breaches.

Defining AI along a spectrum of capabilities

While genAI has driven most of the hype around AI since late 2022, AI capabilities sit along a spectrum, with each advancing level offering different capabilities and/or deeper intelligence. At its simplest, core AI is the application of intelligence in machines. Machine learning extends this by working with larger datasets. Generative AI goes further to enable content creation without the need for defined input parameters.

AI capabilities explained

	Core AI		
		Machine learning	Generative AI
Simple definition	The use and application of intelligence by computers and computer systems.	AI that uses computer systems to detect and learn from statistical patterns, using vast quantities of data.	The use of AI (often via neural networks) to create content. Learned patterns reduce the need for input parameters over time.
Practical meaning	Applications vary hugely – across military, commerce and academia, for example.	The ability to run intelligence on a wider range of applications with larger data trails (e.g. drug trials or demographic patterns).	Adds creation to the mix. This can manifest in multiple forms including text, audio, video, mathematical syntax and shapes.
Development timeline	1940s to present	1950s to present	2000s to present
Examples	Generally available (academia and private sector)	IBM Watson, DeepMind	ChatGPT (OpenAI), Ernie Bot (Baidu), Gemini (Google)

Source: GSMA Intelligence

A broad, competitive landscape for AI

- AI is a product of data science. As such, it involves a broad competitive landscape.
- The bedrock of AI capabilities is in enhanced computing power, and specifically the advent of GPUs and NPUs – an evolution from CPUs that has powered computers and smartphones for 30 years. Nvidia has established a strong position – but competition is extensive from the likes of Intel, AMD and Huawei (in China).
- LLM developers and hyperscalers play a complementary role developing the core algorithms and providing the workhouse power to train and use (or inference) them in practice. Enterprise IT can also be a producer and consumer of AI.
- Telcos will interface with all of these, depending on the AI use case.

AI landscape

LLM developers

- Develop the algorithms that underpin LLMs
- Offer tailored software solutions for business operations and data management
- AI is used to enhance performance and efficiency for edge and networking solutions
- Key players include OpenAI, Anthropic, numerous start-ups, Hugging Face (open-source platform of trained LLMs)

Operators and LLM developers can collaborate by jointly developing, sharing data, training and setting up feedback loops to create bespoke, compliant LLM solutions for the telecoms industry.

Silicon and chipset providers

GPU and CPU

- Graphical and computer processing chips required for AI and ML computing
- Deployed in datacentres and other advanced computing environments (e.g. gaming)
- Key players include AMD, Intel, Nvidia and in-house designs (e.g. Huawei)

System-on-chip (SoC)

- Chipsets with pre-programmed software
- Deployed in advanced computing environments for specific form factors (e.g. cars or robotics)
- Key players include AMD, Broadcom, Intel, Qualcomm, Nvidia

Operators can use advanced silicon and chipsets to boost AI capabilities for efficient network operations and advanced data processing.

Enterprise IT and edge providers

- Provide edge and other IT infrastructure for enterprise/B2B customers:
 - localised data centres and edge servers near or on-premises
 - networking and routing
 - enterprise IT software
- AI is used to optimise edge and networking solutions
- Key players include Dell, Fujitsu, HPE, Marvell, Oracle, Supermicro

Operators can use IT infrastructure and edge solutions to enhance their network efficiency, reliability and customer-service capabilities.

Hyperscalers

- Operate global data centres that house server infrastructure used to power AI and other accelerated compute workloads
- Support public, private and hybrid cloud environments
- Use a mixed AI chip sourcing strategy, combining external procurement (e.g. Microsoft using Nvidia) with in-house development and iteration
- Key players include Alibaba, AWS, Google, Microsoft

Operators can collaborate with hyperscalers for scalable computing power and enhanced AI support.

Recent developments: operator ecosystem

Key launches and announcements in Q2 2024

<p><u>KT unveils three AI power saving technologies</u></p> <p>KT announced three AI technologies to reduce power consumption: Equipment Temperature Management, Energy Saving Orchestrator, and Server Power Supply Optimization.</p>	<p><u>Airtel and Google collaborate on cloud and genAI</u></p> <p>The companies will work together on sales efforts, pairing Airtel connectivity and Google AI assets such as an Airtel IoT solution and managed services.</p>	<p><u>Verizon unveils new AI tools to transform customer experience</u></p> <p>The operator is deploying human-aided genAI to simplify user experiences, aiming to reduce “cognitive load on store and customer-service partners”.</p>	<p><u>Swisscom launches Swiss AI Platform</u></p> <p>Building on previously announced work with Nvidia, Swisscom announced its Swiss AI Platform, offering access to AI and genAI solutions.</p>
<p>April 2024</p>	<p>May 2024</p>	<p>May 2024</p>	<p>May 2024</p>
<p><u>Ooredoo Group pioneers AI revolution in MENA region with Nvidia collaboration</u></p> <p>Ooredoo has become an Nvidia Cloud Partner, developing an AI-ready platform for users in Qatar, Algeria, Oman, Tunisia, Kuwait and the Maldives.</p>	<p><u>Global Telco AI Alliance (GTAA) founders sign JV agreement</u></p> <p>At the DTW24 – Ignite event, e&, SK Telecom, Deutsche Telekom Singtel and SoftBank signed an agreement to co-develop a multilingual, telco LLM.</p>	<p><u>Singtel and Hitachi Digital partner to accelerate industrial AI solutions</u></p> <p>Singtel’s Paragon platform for 5G, edge and cloud orchestration will be paired with Hitachi’s AI expertise. It includes a pilot at a US factory for Industry 4.0 use cases.</p>	<p><u>KDDI starts discussions to establish the largest AI data centre in Asia</u></p> <p>Working with Datasection, Supermicro, Sharp and Nvidia, KDDI will work to build an AI data centre, supplying construction, operations and infrastructure support.</p>
<p>June 2024</p>	<p>June 2024</p>	<p>June 2024</p>	<p>June 2024</p>

Source: GSMA Intelligence, based on company announcements

Recent developments: supplier and tech ecosystem

Key launches and announcements in Q2 2024

<p><u>Microsoft lightens up on AI with Phi-3 launch</u></p> <p>The company's small language model (SLM) promises to outperform models of larger sizes, making it suited to resource-constrained environments, and low-latency or cost-constrained use cases.</p>	<p><u>AWS commits to continued data centre investment in key markets</u></p> <p>In its Q1 earnings, Amazon noted multi-billion dollar AWS data centre investments in the US, Saudi Arabia and Mexico – part of a \$150 billion investment over the next 15 years.</p>	<p><u>Nokia launches Industry 4.0 edge compute and AI solution</u></p> <p>The Finnish vendor's MX Grid solution builds on its MX Industrial Edge platform and private wireless assets to support AI/ML data processing at the network edge for enterprise OT use cases.</p>	<p><u>UK, US join forces on AI development</u></p> <p>The two governments announced an MoU to partner on the safe development and use of AI, including joint testing of public models, following a previous deal focused on AI safety and security.</p>
<p>April 2024</p>	<p>April 2024</p>	<p>April 2024</p>	<p>April 2024</p>
<p><u>Vodafone supercharging customer experience with Microsoft's genAI tools</u></p> <p>Building on a partnership signed in early 2024, the telco provided an update on a multilingual virtual assistant improvement and 50% boost to customer-care enquiry resolution.</p>	<p><u>EU AI Act delivers first-of-its-kind rules on AI development</u></p> <p>Approved by the Parliament in March, the EU Council sign-off harmonises AI regulation with a risk-based approach that allows high-risk systems with strict limits, and bans use cases where risk is deemed unacceptable.</p>	<p><u>TM Forum releases L4 Autonomous Networks Blueprint</u></p> <p>Supported by an array of operators and network vendors, the industry group outlined the capabilities of highly autonomous networks and a pathway for deploying them.</p>	<p><u>Apple Intelligence puts genAI at core of iPhone, iPad and Mac</u></p> <p>With its new operating systems, Apple will integrate ChatGPT into its hardware. It later noted that AI features would be withheld in Europe to ensure DMA compliance.</p>
<p>May 2024</p>	<p>May 2024</p>	<p>June 2024</p>	<p>June 2024</p>

Source: GSMA Intelligence, based on company announcements

Learnings from DTW24 – Ignite

TM Forum's DTW24 – Ignite took place in Copenhagen in June 2024. There were around 3,900 attendees, including 1,000 from operators, to engage in the organisation's traditional focus areas – digital transformation, automation, operations and billing. AI took centre stage.

AI's role in the agenda

The TM Forum identifies three 'missions': composable IT & ecosystems, autonomous network operations, and data & AI innovation. A focus on AI was therefore evident across two of these.

DTW24 themes

Official show themes included AI-native DNA, implementing AI, autonomous networks, modernising IT, and platforms & partners. Beyond the 40% with a direct AI focus, the role of AI in the others was also clear, with an IT and networks focus aligning well with AI-led operations.

DTW24 awards

AI was an overarching theme across the Excellence Awards, including Bell Canada's use of AI for API management, and China Mobile leveraging TM Forum's AIOps Framework (with Huawei) and genAI for autonomous network operations (with Nokia). Beyond operations, e&'s use of data and AI to drive new revenue streams addressed the pressing need to monetise network assets.

Implications for the industry

Trade shows and industry events are natural triggers for new product and service messaging. DTW24 was no different. Helping characterise the show's AI message were the following:

[Matrixx demos AI-accelerated digital revenue streams](#)

[Netcracker accelerates digitisation in satellite communications](#)

[Nokia and Google Cloud collaborate on telco APIs](#)

[Tecnotree and TCL partner on 5G-led genAI BSS solutions](#)

[Telefónica Tech and IBM collaborate on AI for enterprises](#)

The biggest news from DTW24 was the signing of the Global Telco AI Alliance joint venture. While there was plenty of activity around AI for network operations, AI in support of service creation and revenue generation was front and centre. This aligns well with telcos' interest in leveraging AI to grow their business. It remains to be seen if and how these solutions are put to use, and what they can deliver.

Learnings from MWC Shanghai 2024

MWC Shanghai 2024 was held in June attracting nearly 40,000 visitors and 6,500 companies. The event is a sister event to MWC Barcelona, extending the platform but viewing key industry trends and innovation through a Chinese and (more broadly) Asian lens. AI was a prevalent theme both in its own right and in its influence across other topics – notably, 5G-Advanced networks and energy.

AI's role in the agenda

The topic of AI cut across the event. Discussions and demonstrations of AI covered the rapid innovation cycle for the technology itself, and expectations for its influence on telcos' operating models and cost structures.

Tech innovation

China is a global leader in many advanced technologies (such as 5G and quantum computing). This is also the case for AI. Discussions focused on LLM development and applications for specific scenarios in network operations and customer services, evidenced by announcements from several Asian operators and partners.

Economic impact

The agenda also featured the notion of an 'AI economy', the telco role in shaping that, and implications for the labour market, GDP growth and geopolitics.

Implications for the industry

Announcements and demos at MWC Shanghai underline the advanced position of the Chinese operators and other Asian telcos – particularly those from South Korea and Singapore. These included the following:

[Chinese operators speak at the vanguard of AI](#)

[Singtel overhauls customer contact centres with AI](#)

[Handset makers use AI to bolster security](#)

Near-term implications include network cost structures (improved energy efficiencies, reduced labour hours), customer engagement and security. It was also telling that CEOs from the three largest Chinese operators indicated AI was now a fundamental driving force in their corporate strategy. Particularly notable was the truth that, despite US sanctions on IP exports to China (notably from Nvidia), the domestic semiconductor sector has obviated foreign reliance at breathtaking pace.

Filtering out the noise to identify the key themes

AI has dominated telecoms newsflow thanks to activity from operators, network vendors, cloud providers and the broader telecoms ecosystem. Across the announcements, partnerships and product announcements, several key themes have emerged.

<p>Telco AI is much more than genAI, with networks a focus</p>	<ul style="list-style-type: none"> • Operators are taking an active role in genAI deployment, tapping it for customer care, new services and telco-specific LLMs. • Across energy saving, network operations support and GPU rental services, the breadth of activity stretches well beyond genAI.
<p>Cloud co-opetition is alive and well in AI</p>	<ul style="list-style-type: none"> • Cloud providers have long been positioned as telco ‘frenemies’ – competing for customers on some fronts but also serving as valuable partners for internal and customer-facing use cases. • This same dynamic is seen in telco AI activity. While some telcos are developing their own cloud capabilities to support AI, many are also working with cloud incumbents to leverage their AI assets and expertise.
<p>AI is supporting in-market cloud activity, with telcos in the mix</p>	<ul style="list-style-type: none"> • Some AI deployments involve cloud computing assets operated only within an organisation’s local market, ensuring personal data is processed in line with national legislation requirements. • Potentially sensitive AI applications may drive demand for in-market solutions. Market expansion by major cloud players and cloud deployments by telcos in their domestic markets are one result.
<p>AI-enabled user devices are coming (maybe)</p>	<ul style="list-style-type: none"> • While the cloud currently plays a dominant role in AI, running AI applications on user devices – including PCs smartphones and IoT devices – promises security, latency and even energy-efficiency benefits. • Slimmed-down LLMs combined with powerful device silicon can deliver on these promises. Users must be convinced of the advantages, while manufacturers must be sure the AI support complies with regulation.
<p>Regulation’s impact on AI is moving from theoretical to broader policy and regulation</p>	<ul style="list-style-type: none"> • Business decisions are generally motivated by operational efficiencies, user experience and revenue improvement, or regulatory compliance. AI deployment decisions are no different. • Governments around the world have a clear interest in ensuring robust AI innovation but also safe and responsible deployment within the framework of broader policy and regulation.

01

Executive summary
Page 5

02

Market context
Page 8

03

Leader profiles: telco AI in action
Page 17

04

Deep dive: core AI versus genAI
Page 22

05

Deep dive: open versus
proprietary LLMs
Page 28

03

Leader profiles:
telco AI in action

AI in action: e& GenAI Market Research and Survey Tool

e& UAE is adopting AI and data-driven solutions. The company plans to leverage AI for greater personalisation, network optimisation and service innovation, strengthening its position as a future-proof and customer-centric organisation.

Challenge

e&'s business and consumer departments frequently conduct market research to understand user behaviour, Net Promoter Score and trends. Achieving high response rates has traditionally relied on voice calls rather than online, SMS or other methods. This required an external agency, which was time-consuming and could cause delays. It took a minimum of 10 days and incurred significant costs to perform the survey and market research, along with potential security issues. To address this, e& developed a trained virtual agent, creating a robust market research and survey tool.

Solution

The GenAI Market Research and Survey Tool has been built to enhance efficiency and effectiveness. The goal was to streamline survey operations, reduce reliance on an external agency, and cut the time and cost associated with traditional voice-call surveys and market research.

The solution features a virtual agent capable of autonomously conducting voice-call surveys. Business users can create surveys by describing them, with the system automating the process. This stretches from end to end, including creating market research survey questions using genAI LLM, automated testing, conducting the survey or market research via a virtual agent, and producing results and insight analysis using genAI.

Key capabilities and technology used

- **Multilingual surveys** – The solution generates full survey questions and market research in multiple languages. The virtual agent is a well-trained AI system capable of efficiently handling voice interactions.
- **Survey management and data integration** – GenAI automates survey execution and management. Customer data is gathered from data warehouses are used to store and retrieve customer data and responses.
- **A unified marketing platform** – Contacts are pushed to Pega, an AI-based platform for managing customer interactions and data.

AI used: Gen AI GPT-T4, Virtual Agent Conversational application, Azure Speech suite, NLP, Avaya Telephone System, Pega, Virtual Surveyor, Java orchestrator.

AI in action: e& GenAI Market Research and Survey Tool (continued)

Approach and learnings

e& has focused on in-house product evolution along with the development of organisational capacity and ensuring the ethical use of AI.

- **The ecosystem** – The operator developed the tool in-house using existing technologies. To build this use case, the only newly introduced technology was genAI. Partnerships have been established with Pega, Cloudera, Azure OpenAI and Azure Speech.
- **Building capacity through upskilling** – To maximise the potential of the new AI framework, e& has invested in upskilling. Employees received training to manage and use AI effectively. They are competent in building surveys using the technology provided and interpreting the data generated.
- **Managing data** – Strict protocols were established to ensure data integrity and accuracy, essential for making informed business decisions based on survey results.
- **Safety and risk mitigation** – Stringent data privacy and security measures safeguard customer information accessed and processed by AI systems. All components are hosted on-premises, with the GPT model operating within e&'s Microsoft Private Cloud in the UAE. Regular audits and system updates are conducted to mitigate risks associated with technology failures or data breaches.

The implementation began with pilot testing to refine genAI capabilities before full-scale deployment. Technical challenges included integrating AI with existing systems, ensuring seamless data flows between the data warehouse, Pega platform and voice virtual agent, and achieving accurate natural language processing in creating surveys.

Through the genAI capabilities, the operator achieved a significant reduction in staff hours worked for survey operations, streamlined processes and enhanced operational efficiency.

The tool demonstrated a high response rate (an increase of 58% compared to online surveys), underscoring its effectiveness in engaging respondents.

The solution could be monetised as a platform-as-a-service across industries. This offers a new revenue-generating opportunity for the operator.

90%

The use of automated genAI capabilities enabled a saving of 90% in staff hours worked

99.3%

Actionable insights are generated 99.3% faster than traditional methods.

AI in action: Telefónica's Kernel 2.0

Telefónica's AI strategy focuses on responsible and ethical implementation of the technology across its operations, guided by its 'Principles of AI' and a comprehensive governance framework. The company has deployed more than 500 AI systems and is committed to continuous improvement and ethical AI development.

Challenge

In focusing on a data-centric strategy during its digital transformation, the operator faced several challenges, including data silos, manual data privacy management, lengthy time to market for digital services, and a lack of standardisation in the use of AI.

To address these challenges, it prioritised maximum use of data by developing a standardised platform, Kernel, in 2017. With genAI, Kernel is now evolving to Kernel 2.0.

Solution

- Kernel serves as the cornerstone to the operator's digitisation plan. It standardises data access, automates privacy and security, and facilitates the seamless creation of digital products and services, through a unique platform, enabling development of digital ecosystems. By integrating genAI capabilities, the goal is to make the most of the potential of LLMs/SLMs developed by the industry in a safe and ethical way.
- The solution facilitates seamless integration of data, APIs and new AI models, and frees developers from the burden of privacy management, allowing them to focus solely on improving features and capabilities. GenAI in Kernel will drive a complete business transformation and apply AI at scale across all functions of the organisation.

Key capabilities and technology used

- **Unified data access** – The solution consolidates Telefónica's databases, providing unified and standardised data sets across the enterprise. The central repository supports analytics, AI model training and decision-making, facilitating AI innovation.
- **Technological enablers** – The solution integrates a unified single digital identity and open APIs for seamless third-party integration (OAuth 2.0, OIDC, OpenAPI). It includes software capabilities (payments, security features, tokenisation, AI tools), supports hardware (routers, set-top boxes) and centralises catalogues to unify service offerings across its ecosystem.

AI in action: Telefónica's Kernel 2.0 (continued)

Approach and learnings

Harnessing genAI capabilities, Telefónica has focused on developing in-house capacity and ecosystem development, while prioritising privacy by design. Kernel 2.0 is enabling the operator's digital capabilities – from complex telecoms systems to new, customer-centric digital solutions.

- **Building capacity** – Kernel is an in-house development that is 100% cloud based, leveraging a combination of internal, open-source and commercial software. Telefónica's data scientists build the genAI capabilities, providing complete control of the solution.
- **Integrating diverse in-house expertise** – Coordinating a number of multidisciplinary teams across Telefónica was key to development. Under its privacy 'pillar', designing a common language (digital privacy framework) between lawyers, UX designers, software, privacy and security engineers and architects was crucial. For the APIs and identity pillar, collaboration between the security and IT areas was important. Effective teamwork among data scientists, engineers and regulation experts was essential for the AI and data pillar.
- **The ecosystem** – To build a strong technological foundation, the operator led the development together with technology partnerships. Microsoft is the main provider of cloud services and AI enablers, while the solution relies on third-party LLMs.
- **Privacy by design and data security** – Telefónica ensures all AI models and applications built on the platform adhere to strict privacy standards. By centralising privacy management, Kernel minimises the risk of inadvertent or negligent access to personal data, safeguarding customer privacy and complying with global regulatory frameworks. Robust security protocols and governance mechanisms ensure the security of sensitive data.



New services

The integration of third-party LLMs/SLMs with operator data and APIs via centralised control has enabled diverse use cases, including Kernel's creation of trusted copilots for natural language access to operator data, enhancing decision-making. These applications leverage retrieval augmented generation (RAG) technology such as genAI models embedded in Kernel and trained on internal data. In addition, other use cases are advancing autonomous and trustworthy agents (such as software programmes and systems) by using Kernel APIs based on the same architecture.



AI capabilities and innovation

By using AI, Kernel optimises the customer experience and maximises business value through AI as a service. The integration of genAI is driving a complete business transformation, applying AI at scale across all functions of the organisation.

01 Executive summary
Page 5

02 Market context
Page 8

03 Leader profiles: telco AI in action
Page 17

04 Deep dive: core AI versus genAI
Page 22

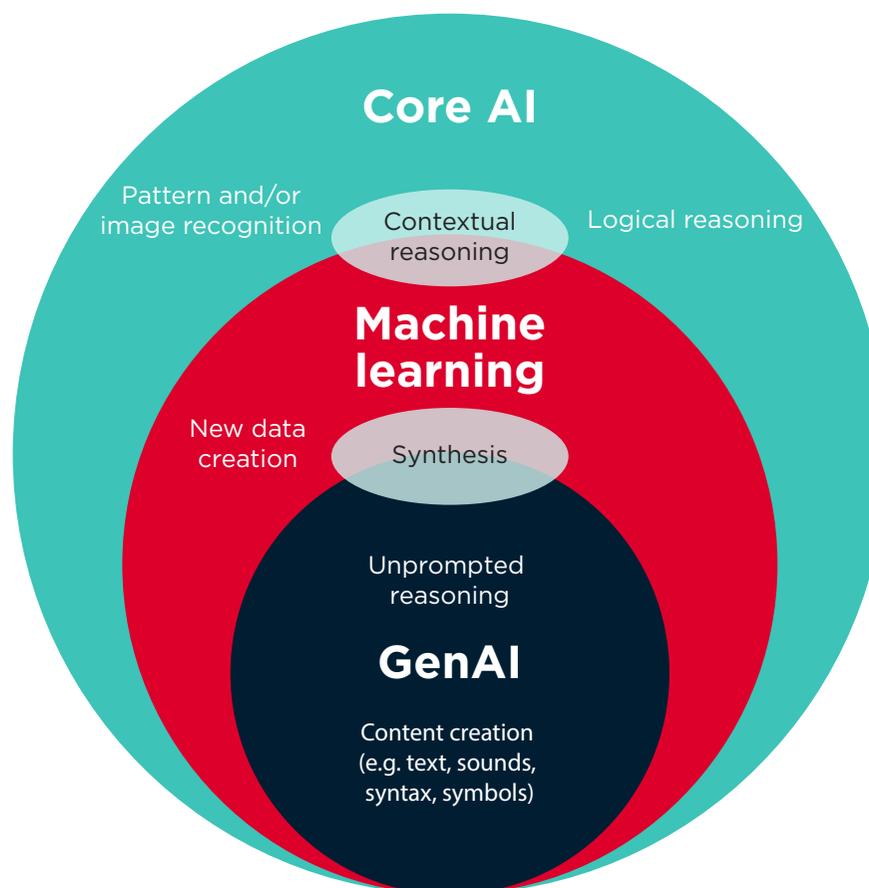
05 Deep dive: open versus
proprietary LLMs
Page 28

04 Deep dive: core AI versus genAI

Core AI to genAI: deeper intelligence means deeper capabilities

- **A series of overlapping technologies** – AI is effectively a series of overlapping technologies (rather than one distinct type of software or code), differentiated based on the type of intelligence and pre-emptive action/creation that can come from its use. While genAI has grabbed most of the attention of late, many companies (including telcos) can benefit from core AI.
- **GenAI is a complement not a replacement** – GenAI is more a complement to (rather than replacement for) core AI or machine learning, adding creation and synthesis capabilities. This could, for example, provide network engineers with recommendations on which sites to prioritise for upgrades, or offer customer services suggestions for resolving a given problem.
- **Strategic investment should drive thinking** – The key is understanding where genAI is needed (and therefore merits investment) versus where it is not (a simpler off-the-shelf form of AI could suffice).

AI and example capabilities

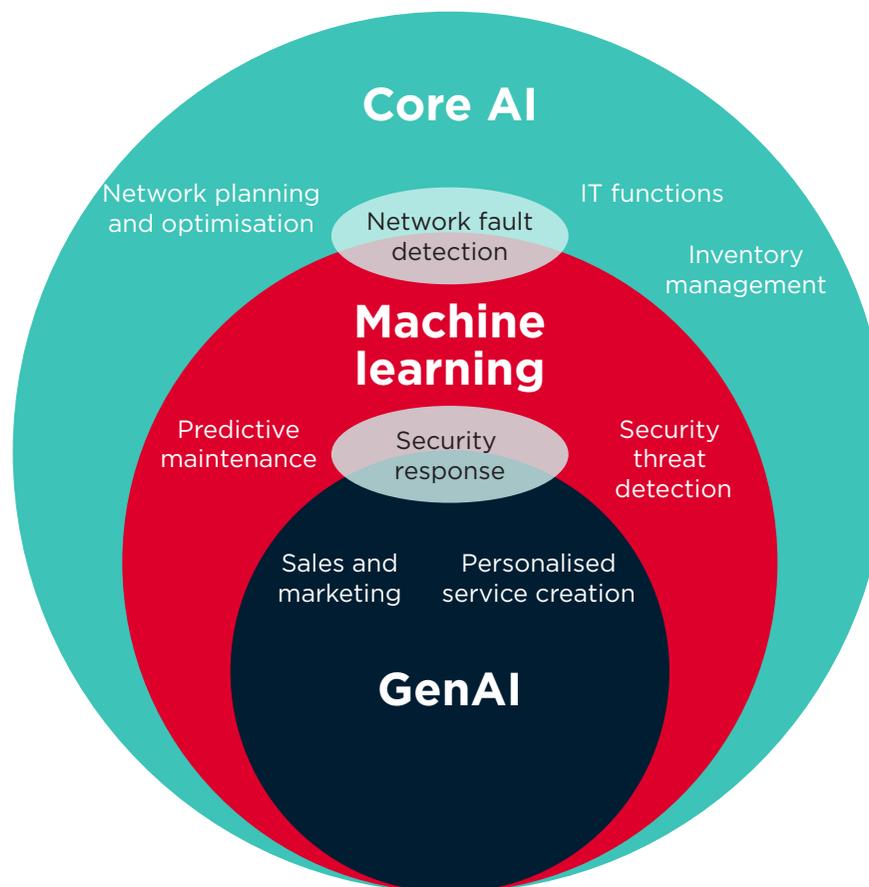


Source: GSMA Intelligence

AI in telecoms: the right tech for the right use case

- **Move early and iterate** – It is easy to overlook the power of core AI amid the ChatGPT-inspired genAI activity. AI is already being used by many telcos in network operations, energy optimisation, customer call centres for automatic language recognition, and retail operations.
- **Services are next** – Early deployments in telecoms are concentrating on easy wins such as network fault detection and automating more of the functions used in customer-care centres. As benefits start to accrue, genAI will likely be used more widely. Examples include creating personalised services, segmenting customers into new demographics, and managing network energy.
- **The right tech for the use case** – This is ultimately a generational technology that will be phased in over time. The success/value of AI use comes down to selecting the right technology for the right use case. To this end, notions of ‘betting the farm on core AI’ or ‘we’re all in on genAI’ risk obscuring a more nuanced playing field.

Examples of telco-specific use cases with AI



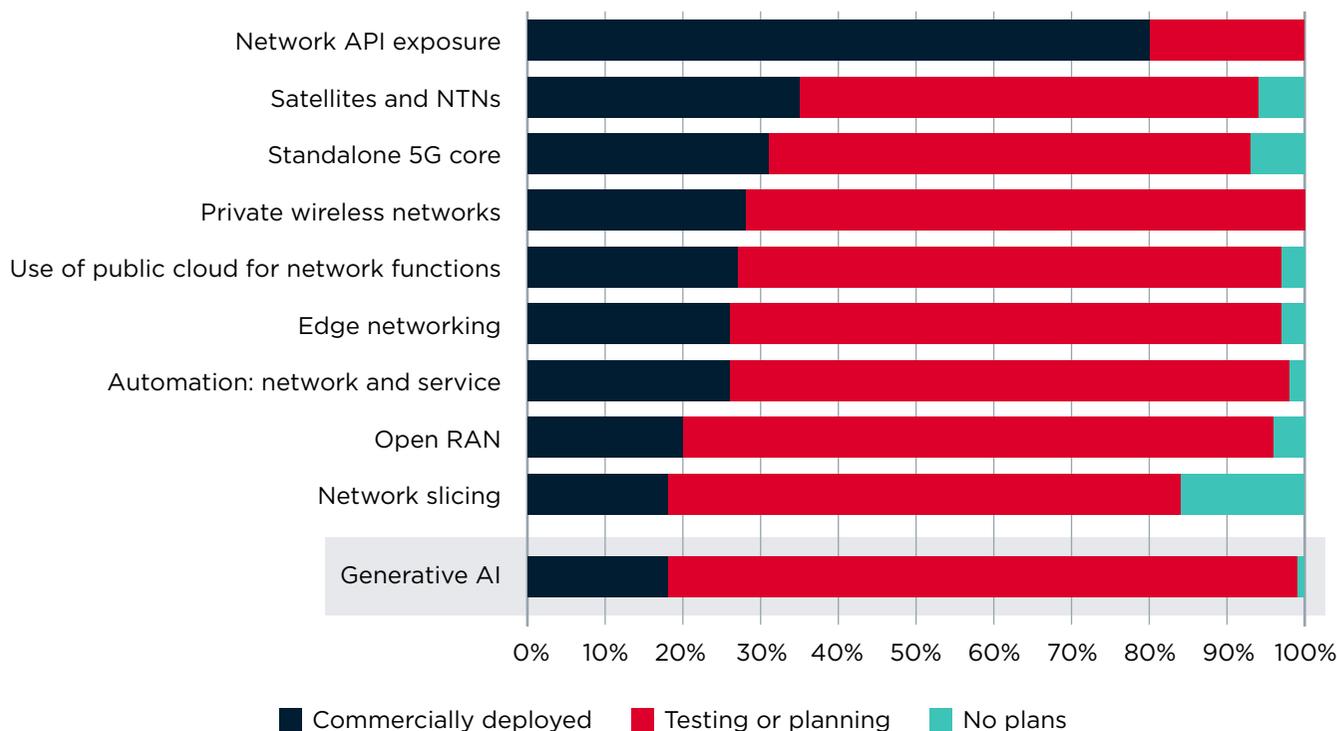
Source: GSMA Intelligence

AI is both a sprint and a marathon

- **Everyone gets it** – AI is the most transformative technological change since the internet, something that is as true for telcos as it is for everyone else.
- **Numbers mask much of the story** – Numbers are one proxy for action, though these only tell part of the story. Even as of mid-2023 – well after ChatGPT’s implications had become clear – only around 20% of operators reported that AI was a part of live operations. The reality is that closer to 100% of operators now likely have AI in active use. The ‘testing or planning’ cohort (see chart) and chorus of CEO voices reflect this.
- **Easy wins and slow burns** – Variation in the industry comes from the breadth and depth of AI use. Most operators will have trialed or deployed AI in some way, such as in network maintenance or customer service. These are easy wins and a ‘sprint’ to get right because of the competitive dividend from satisfied customers and lower costs. The longer-term remodelling of how business is done and operated is more akin to a ‘marathon’, with a joined-up approach between technology, product and strategy teams required.

Where do you stand in the deployment of each technology?

Percentage of operators



Source: GSMA Intelligence Network Transformation Survey 2023

Proving AI's value in telecoms

- **Network is the bedrock** – The network is the natural place to start with AI for operators. GSMA Intelligence survey data supports this, with around two thirds of operators viewing troubleshooting and maintenance as having the biggest impact from genAI.
- **Services are the logical follow-on** – Service creation remains at an early stage but could involve new forms of customer segmentation (rather than traditional age/gender/income models) to better target services where demand lies.
- **Proof points start with P&L** – The return on AI is tough to measure but is likely to start in the profit and loss (P&L). As an example, energy still accounts for 15–20% of opex for operators on average. Reducing power consumption by 20% (something AI is already helping with via sleep states and antenna design) would lower opex by around 4%.
- **Not all operators are the same** – Survey data points to a divergence between large and small operators. Namely, that smaller telcos:
 - have more urgency in addressing pain points from operating smaller networks (such as mitigating traffic load)
 - appear more willing to experiment with AI on a wider range of uses earlier on, such as in customer experience and service creation.

Where telcos think genAI will have the greatest impact on their business

How do you believe generative AI technologies will have the greatest impact on your business? Pick the top two

	All operators	Small cap (<\$500 million in annual revenue)
1	Network troubleshooting and predictive maintenance (66%)	Personalised service creation (46%)
2	Threat detection and mitigation (47%)	Network planning and optimisation (38%)
3	Personalised service creation (28%)	Network troubleshooting and predictive maintenance (31%)

Source: GSMA Intelligence Network Transformation Survey 2023

Core AI and genAI will each play a role

Core AI and genAI use are likely to happen in parallel for many years. There are several pragmatic reasons for this (see below). These include the fact that AI can suffice for many of the tasks required, costs, LLM creation and refinement, and the time taken to train and recruit staff with specialisms in AI/data science/software engineering. Over time, the genAI term may dissolve into AI as the generative functions (and LLMs behind it) become commonplace.

Tech availability

- **AI value proposition** – As much as genAI can create content in the form of words, syntax or code, AI will increasingly underpin automation of network functions and other software-driven tasks. Hype should not distract from valid investments in core AI for years to come.
- **LLMs** – There is also the question of sourcing the right AI LLMs, which takes time but these are essential to develop genAI. The Telco LLM venture (by the Global Telco AI Alliance) is there because it allows resources to be pooled, but many groups will continue to use off-the-shelf AI for years to come.

Costs

- **P&L** – Upfront and ongoing costs (such as licensing) associated with developing genAI are significant. This is not a barrier to eventual genAI use in more of the telco business, but it does limit how fast a transition can take place.
- **Energy** – Hyperscalers account for 1% of global energy use. AI training and inference is likely to drive that up to 1.5–2.0% in the next five years. This implies the need for renewable offsets, directly and via partners (including telcos).

Staff and human capital

- **Oversight and assurance** – Regardless of the benefits genAI brings, most functions in networks and customer-facing roles will continue to require human oversight. This is likely to involve a mix of automation and supervised genAI.
- **Upskilling time** – Training and recruiting staff skilled in data science and AI engineering is a 10–15-year project – a prerequisite for many of the assurance requirements.

01

Executive summary
Page 5

02

Market context
Page 8

03

Leader profiles: telco AI in action
Page 17

04

Deep dive: core AI versus genAI
Page 22

05

Deep dive: open versus
proprietary LLMs
Page 28

05

Deep dive:
open versus
proprietary
LLMs

LLMs are reshaping telcos

- **LLMs have the power to build efficiency** – LLMs represent the front line of the AI revolution, providing unprecedented capabilities for data processing, customer interaction and network management. In the telecoms sector, LLMs enhance customer support by providing 24×7 service and accurate responses. They improve operational efficiency through automation of tasks such as billing and technical support, leading to cost savings. Additionally, LLMs enable personalised services by analysing customer interactions.
- **Co-creation spurs adoption** – With LLMs becoming essential in telecoms, collaborations such as the Global Telco AI Alliance's development of telco LLMs and partnerships between Amdocs, Nvidia and Microsoft Azure have accelerated adoption. These collaborations facilitate the creation of customised LLMs tailored for telecoms applications by combining expertise. For example, Amdocs contributes telecoms knowledge, Nvidia provides advanced AI capabilities, and Microsoft Azure offers robust cloud infrastructure.

Key challenges ahead

- Early telecoms efforts in genAI focus on enhancing customer service with LLMs. However, domain-specific LLMs may be needed for areas such as marketing and network operations.
- Creating a telco-specific LLM requires substantial capital. This is prompting operators to explore cost sharing. For instance, in 2023, SK Telecom invested \$100 million in AI research company Anthropic to develop a telco-focused LLM.
- Achieving a universally applicable LLM across diverse telecoms environments is challenging, with success depending on how well LLMs are refined and adapted for each operator's needs.

What are LLMs?

LLMs are a type of deep-learning algorithm designed to recognise, summarise, translate, predict and generate content using vast datasets.

LLMs are a component of genAI, which is named for its ability to produce a diverse range of outputs, including text, images and audio. While LLMs focus specifically on text generation, genAI encompasses a broader spectrum of content creation capabilities.

Why are they important?

Their importance spans numerous industries, enhancing productivity, efficiency and the quality of services provided. From customer support and content creation to healthcare and education, LLMs have diverse and impactful use cases, driving innovation and improving human-computer interaction across the board.

Open-source versus proprietary LLMs

- Understanding the trade-offs** – The advantage of using open-source LLMs lies in their cost effectiveness, rapid implementation and ability to leverage community-driven innovation and updates. The proprietary option means LLMs can be bespoke to fit a certain need (e.g. network maintenance), but these may cost more to build.
- Operators target local opportunities** – LLMs developed by US tech companies excel in English, but their performance in other languages varies based on the complexity and availability of training data. This creates an opportunity for locally developed, telco-specific LLMs. In Asia Pacific, operators are capitalising on this opportunity by building their own LLMs and genAI services.
- Examples in practice** – Using telecom-specific LLMs trained with local language data can yield superior results. For example, SK Telecom reported that its telco-specific LLM performed around 35% better than general LLMs. In customer-facing use cases, the telecoms LLM delivered more contextually accurate responses, leading to higher customer satisfaction, as highlighted at DTW24

Open source	
Build type	Advantages
Use Utilising existing open-source LLMs without significant modifications.	<ul style="list-style-type: none"> • Cost-effective deployment • Quick implementation • Leverage community innovations
Build and fine-tune Modifying and enhancing open-source LLMs to better suit specific needs.	<ul style="list-style-type: none"> • Tailored solutions • Control over modifications • Encourages local innovation
Examples: LLAMA 2 by Meta, GPT-NeoX by EleutherAI	

Proprietary	
Build type	Advantages
Buy Purchasing proprietary LLMs developed by other companies.	<ul style="list-style-type: none"> • Optimised for specific tasks • Vendor support • High performance
Build and fine-tune for specific need Developing proprietary LLMs and adjusting them to meet specific requirements.	<ul style="list-style-type: none"> • Highly tailored solutions • Complete control over data and security • Competitive advantage
Build and use (for intended purpose) Creating proprietary LLMs from scratch and deploying them within organisation.	<ul style="list-style-type: none"> • Custom, end-to-end solutions • Unique features • Strong brand differentiation
Examples: GPT-4 by OpenAI, PaLM 2 by Google	

Note: Both open source and proprietary LLMs in the telecoms sector benefit significantly from fine-tuning, a process essential for optimising models to specific tasks or datasets.

Source: GSMA Intelligence

Understand and benefit from the trade-offs

The choice between open-source and proprietary LLMs is multifaceted and requires consideration of the pros and cons of each to make an informed decision.

- Consider resources and purpose** – Developments in LLMs highlight the diverging approaches in their use among telcos. Many adopt a hybrid approach, leveraging open-source models for customisable and innovative aspects of projects while relying on proprietary models for critical areas requiring enhanced security, reliability and robust support aligned with business needs. Examples of different approaches include the following:
 - Safaricom in Sub-Saharan Africa uses open-source LLMs such as GPT-3 for AI-driven customer support and content delivery, benefiting from community-driven innovations and cost-effective solutions.
 - Vodafone in Europe employs proprietary voice assistants integrated with LLMs for secure mobile banking and personalised customer support, ensuring optimised performance and meeting stringent regulatory requirements.
 - AT&T in the US uses a hybrid approach for network management and customer service. It customises open-source models for flexibility and innovation while relying on proprietary solutions for critical operations requiring enhanced security, reliability and specialist functionality.
- Regional variation persists** – Operators in the US and EU often adopt open-source LLMs, fostering innovation and compliance with stringent data privacy laws. In contrast, Chinese and other Asian operators focus on proprietary LLMs, supported by government investment and tailored to local needs, enhancing rapid AI development and national data protection. Sub-Saharan African operators rely on open-source models due to resource constraints. They also seek proprietary solutions for high-impact applications through international partnerships.

Open source versus proprietary LLMs: key factors

Key factors	Open-source LLMs	Proprietary LLMs
Cost	✓	✓ ✓
Community support	✓ ✓	✓
Customisation	✓	✓ ✓
Time for development	✓	✓ ✓
Skills	✓ ✓	✓ ✓
Maintenance and support	✓	✓ ✓
Security	✓	✓ ✓

Source: GSMA Intelligence

A basic 'how to' guide

The prerequisites that operators must consider with LLMs

- To effectively use LLMs, operators need robust infrastructure with high-performance computing and scalable cloud services. They must have access to large, relevant datasets and possess technical expertise in AI, machine learning and NLP.
- Ensuring data security and privacy compliance is crucial, along with the capability to integrate LLMs with existing systems. Continuous model maintenance and performance monitoring are essential, as is training staff to adapt to new workflows.
- Partnerships with technology vendors can provide additional support for successful LLM implementation.

Do all genAI solutions need LLMs?

- Not all genAI solutions require LLMs, as genAI encompasses other AI technologies such as computer vision, machine learning algorithms and robotics.
- The use of LLMs depends on specific tasks, focusing on NLP such as text generation, translation and understanding human language patterns.
- Operators leverage LLMs for automating customer support with AI chatbots, optimising network management through predictive analytics, and personalising marketing strategies with customer data analysis.

How are SLMs different to LLMs?

- SLMs are designed to be efficient and require fewer computational resources, making them suitable for simpler language tasks and use on devices with limited power. LLMs are much larger and more complex, trained on vast datasets; they are capable of advanced natural language processing.
- While SLMs are ideal for basic applications such as simple chatbots and text classification, LLMs excel in generating sophisticated text, answering complex questions, and providing deep contextual understanding.
- Telcos use SLMs for customer-support chatbots, automated call routing, and text message automation. In Asia and Africa, where data sources are limited, SLMs are essential for improving multilingual support, voice recognition, and overall customer interaction and network management.

The decision to adopt an open-source or proprietary LLM should hinge on a thorough assessment of project requirements, available resources and associated risks such as security and compliance. Factors including the nature of the solution – whether it is a fintech application or a survey tool – will heavily influence this choice.

It is important to recognise that the optimal decision may differ from one project to another, and within different operational contexts.



GSMA
Intelligence

gsmaintelligence.com

@GSMAi