TM Forum | December 2023

tmforum knowledge

REPORT

generative AI: operators take their first steps

Author:
Mark Newman, Chief Analyst

Editor:
Ian Kemp, Managing Editor

Sponsored by:

Delivering the Digital Future
Netcracker
An NEC Company

contents

tmforum knowledge

REPORT

the big picture

Generative AI (GenAI) technology exploded into the consciousness of both consumers and the business community in November 2022 with the launch of a service called ChatGPT by a little-known company called OpenAI. One year later, perceptions of GenAI have expanded rapidly from an AI-based content creation tool towards a strategic platform that is fast becoming front and center of thinking for almost every communications service provider (CSP) worldwide.

The extent of the impact GenAI is having on the telecoms industry was made clear in a comment by Jonathan Abrahamson, Chief Product and Digital Officer at Deutsche Telekom, in an article for TM Forum Inform: "It's difficult to envisage a world where this doesn't change everything and certainly in the context of what we do as a telecommunications company," he said, referring to the proliferation of GenAI developments just one year after the launch of ChatGPT.

In simple terms GenAI is a type of artificial intelligence (AI) that is capable of creating original content or information, including text, images, audio and video. ChatGPT remains the most widely recognized GenAI-enabled service and the most heavily backed. Microsoft has invested $13 billion in OpenAI, and ChatGPT reportedly now has more than 180 million users and 1.5 billion monthly visitors to its website. But many other enterprise IT tools and services based on GenAI have come to market in the past year resulting in a frenzy of excitement and activity within the global enterprise IT community.

Its arrival did not come as a surprise to everyone. The technologies which GenAI is based on have been in development since the 1960s, and firms such as Google and Microsoft have been working on GenAI-enabled products and solutions for several years.

So how far down the line are CSPs when it comes to adopting GenAI technology? How useful is it? And where does it fit into their business models?

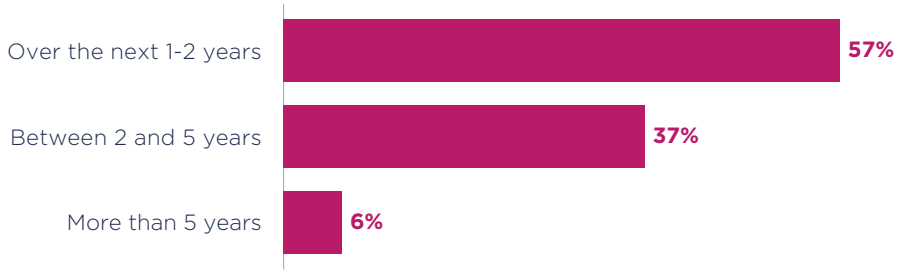
To feed into this report we carried out a comprehensive survey of service providers globally. The charts in the report are based on data from 104 senior-level respondents from 73 operators, just over half of whom are in technology roles.

> **"It's difficult to envisage a world where this doesn't change everything."**
>
> **Jonathan Abrahamson, Deutsche Telekom**

When asked whether their organization is in a good place to exploit the potential of AI and machine learning, 79% said yes. But that still leaves one fifth who think they are not, a function perhaps of the fact that the technology is still in the early stages of commercial development, and that organizations are yet to reshape and take on the levels of skills needed to fully take advantage of these technologies.

But the potential speed and impact of developments may not give them much time to adapt. When we asked when executives think GenAI and large language models (LLMs) – a type of GenAI that can understand and generate human-like language – are likely to have a significant impact on business, nearly 60% said over the next one-to-two years and a further 37% between two and five years.

When do you expect that GenAI / LLMs will have a significant impact on your business?



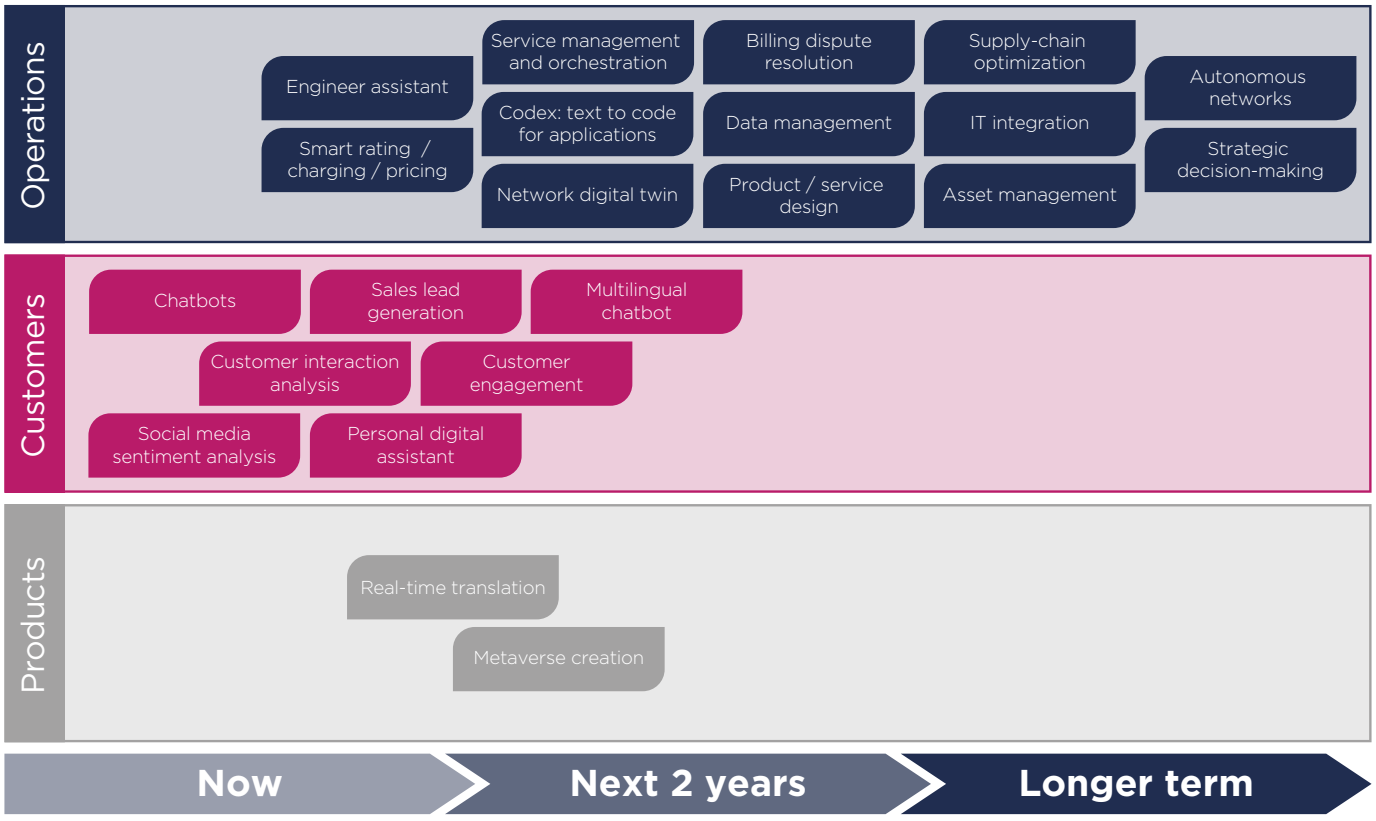
TM Forum, 2023

The graphic below, created by TM Forum and Bain & Company, shows some of the current and longer-term applications for GenAI. We explore early and future use cases in detail in section 5.

Read this report to understand:

- Some of the different technologies and definitions behind the GenAI juggernaut
- Who is doing what in the GenAI value chain
- How CSPs are preparing themselves to implement GenAI
- Early experimentation and GenAI use cases
- The key challenges to developing and adopting GenAI.

Generative AI will be introduced in phases



TM Forum, 2023

tmforum knowledge

REPORT

section 1:

key GenAI technologies and concepts

Generative AI (GenAI) refers to any machine learning model capable of dynamically creating output after it has been trained. The algorithms create new content based on patterns learned from existing data.

GenAI works by processing huge volumes of data to find patterns and determine the best possible response to a question or situation, which it then generates as an output. By feeding the AI immense amounts of data it is able to develop an understanding of correlations and patterns within the data.

Whereas conventional approaches to machine learning require data scientists to develop artificial intelligence from scratch, GenAI involves the use of foundational models – deep-learning neural networks – as a starting point to develop models that power new applications more quickly and cost-effectively (see definitions box on p.9).

Large language models (LLMs) are a type of GenAI that can understand and generate human-like language, with OpenAI's ChatGPT (now at version 4) the most well-known example. LLMs are foundational models specifically trained on massive data sets. They are huge, with billions or even trillions of parameters allowing them to learn complex language patterns and perform tasks that would be impossible for smaller models.

There are many LLMs in existence today and, increasingly, these are designed to serve specific communities or functions. For example, StarCoder and StarCoderBase are LLMs designed specifically for code trained using data from GitHub, while BloombergGPT is an LLM for the financial sector trained on an extensive archive of financial data.

Foundation models are like Swiss army knives which can be used for multiple purposes. Once a model has been developed, any developer can build an application on top of it to create content. The main application of OpenAI's GPT-3 and GPT-4, for example, is the ChatGPT chatbot, but thousands of businesses across the world are now working on their own applications using these LLMs.

What can GenAI do?

The fact that GenAI has a deliverable – an output – makes it different from other types of AI. GenAI can be applied to any content – be it words, images, audio, video or code. This content may be freely available on the internet or proprietary information owned by an organization.

While there is understandable concern about the long-term impact of GenAI – and of AI more broadly – on jobs, it will be used in the short-to-medium term as a tool to help individuals and teams carry out their work more efficiently. Chatbots are deployed as an interface to enable users to interact with GenAI models, and copilots are conversational interfaces that use LLMs to support users in various tasks and decision-making processes. Microsoft 365 Copilot, for example, is a tool, developed using GPT-4 LLMs, that helps people use the company's applications to carry out tasks such as writing documents and summarizing emails.

> **GenAI will be used in the short-to-medium term as a tool to help individuals and teams carry out their work more efficiently.**

The process of experimenting with, and adopting, GenAI for specific use cases is very different from what enterprises are used to with previous models or versions of AI. With GenAI, outputs can be created immediately, but the challenge then is to decide how much fine-tuning is needed to improve the results and what to do with the output.

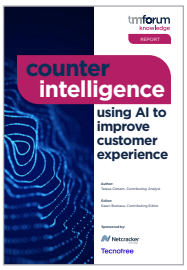
There are also a number of issues and risk factors associated with GenAI – we explore these in more detail in chapter 6 – which need to be addressed, or at least taken into consideration before output content is used. These include hallucinations – inaccurate or inappropriate responses to inputs. Furthermore, many companies are finding that the real benefits of GenAI come in combination with other technologies and types of AI. One large US telco from our survey, for example, has combined GenAI with conversational and enterprise search to make it easier for people to navigate its website.

GenAI use cases

For any enterprise or service provider seeking to exploit the potential of GenAI there is a balance to achieve between pursuing those use cases which are gaining momentum in the business world generally and finding the right internal deployment and democratization strategy to optimize opportunities for innovation. Customer operations represents – by some considerable margin – the family of use cases that is generating the most interest. Interactions between customers and different touchpoints – customer service agents, chatbots, websites and social media – are generating a huge volume of unstructured data, which through the use of GenAI can be converted into knowledge that drives better engagement and opportunities for cross-sell, upsell and contract renewal.

CSPs are already beginning to use generative AI to augment the jobs that human customer service agents perform. Our recent report Counter intelligence: using AI to improve customer experience explains how operators are using AI in general to improve customer experience (CX) and highlights some pitfalls of generative AI.

Read the report to find out more:

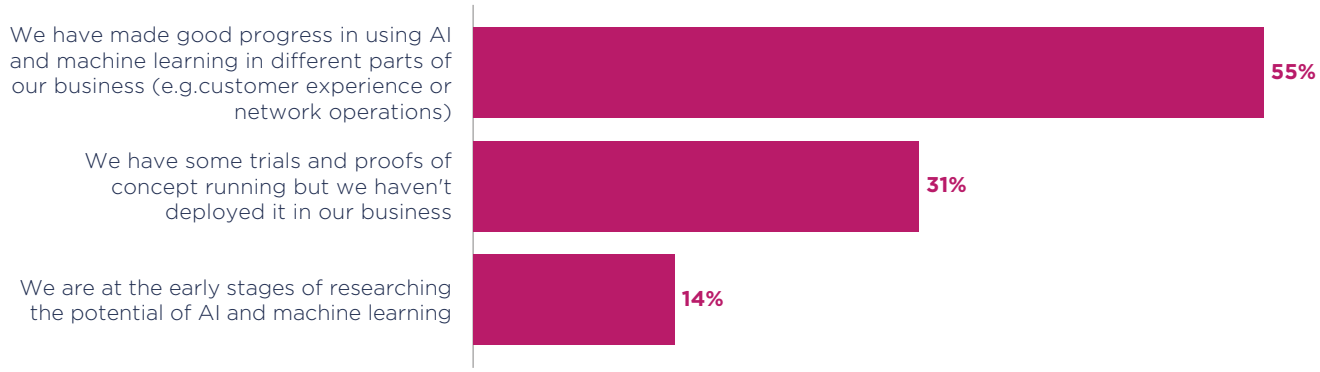


Other broad use cases, and families of use cases, include:

- Sales and marketing
- Network operations
- Fraud reduction
- Product innovation
- Software engineering.

Most of these categories are common to all businesses, with network operations the only category that is specific to telecoms operators (more on specific GenAI use cases in section 5). This is important in the context of the approach that operators take to choosing LLMs and the extent to which they buy off-the-shelf LLMs rather than ones that need to be trained or fine-tuned.

How effectively is your organization exploiting AI and machine learning?



TM Forum, 2023

> **"There is a significant amount of plumbing and orchestration that you have to build around the models to make them work in a telco context."**
>
> **Jonathan Abrahamson, Deutsche Telekom**

Some telco-specific developments are likely to come through partnerships. In July Deutsche Telekom, e&, Singtel and SK Telecom announced the Global Telco AI Alliance to co-develop a platform for operator-specific AI services. Deutsche Telekom and SKT are already working together to develop an LLM for digital assistants in customer service, adapting existing LLMs to learn about telco-specific data.

"[There is a] not insignificant amount of plumbing and orchestration that you have to build around the models to make them work for our [telco] context and...use cases," Deutsche Telekom's Abrahamson told TM Forum. "What became very clear for us is to get really good with this...and to use it at massive scale we need to tune the model to make it work for our use cases."

Amdocs, Microsoft and Nvidia also announced a partnership in November to create LLMs for the telecoms and media industries.

GenAI in telecoms

CSPs have been exploring how and where to adopt AI in their businesses for the past five years or so. In our survey, 55% of respondents said they had made good progress in using AI and machine learning in different parts of their business. But that leaves 45% which are either still at trial stages or just researching their potential (see chart above).

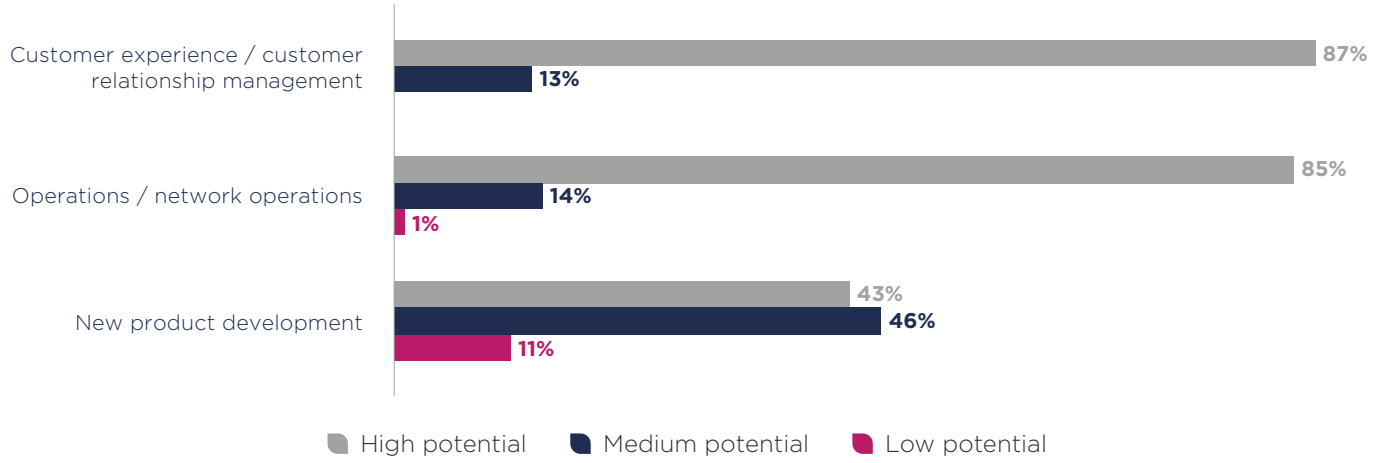
As we saw in The Big Picture, four out of five respondents say their organizations are in a good position to exploit the future potential of AI and machine learning. Despite this confidence expressed by many operator executives, the reality is that the use of AI in CSP businesses remains relatively limited and confined to specific use cases, principally in network operations and customer-facing functions.

Indeed, presented with those use case options as well as product development, a full 87% said AI / machine learning would have a positive impact on customer experience or customer relationship management and 85% on (network) operations (see chart on the next page).

The arrival of GenAI has given a sense of urgency to CSPs in terms

> **"The arrival of GenAI has given a sense of urgency to CSPs to leverage AI.**

Where does AI / machine learning have the most potential to positively impact the telecoms business?



TM Forum, 2023

of how they leverage AI. It comes at a time when operators are desperately seeking inspiration to turn their fortunes around – both in terms of finding new sources of revenue and generating efficiencies. There is a willingness from CSPs to place big bets on GenAI. But at the same time there is an awareness that with the landscape changing so quickly in terms of its technology and players, CSPs need to be able to change direction quickly and easily too.

Research conducted by telecoms, media and technology consultancy Altman Solon on behalf of AWS indicates that telcos will increase their expenditure on AI by as much as six times over the next two years. Almost half of respondents to a survey among 100 CSPs conducted by the consultancy anticipate that GenAI spend will account for between 2% and 6% of total technology spend in two years, up from less than 1% today.

However, it is much less clear where that investment will take place. This will be a function of which approach to building, training and fine-tuning LLMs CSPs take and the commercial relationships with their technology partners. For example, CSPs will be wary of committing to pricing models which scale with usage because of the risk of costs spiraling out of control.

The box on the next page provides definitions of some of the key technologies and concepts around GenAI that we have touched on in this section and cover in subsequent chapters. In the next section we look at where operators fit into the GenAI value chain.

Key technologies and concepts around GenAI

Foundation models

Foundation models – sometimes known as general-purpose AI – are a form of GenAI trained on massive data sets. They form the basis of many applications including ChatGPT – built on the GPT-3.5 and GPT-4 foundation models – and Microsoft's Bing Chat.

Rather than develop artificial intelligence (AI) from scratch, data scientists use a foundation model as a starting point to develop machine learning models that power new applications more quickly and cost-effectively. The term foundation model was coined by researchers to describe machine learning models trained on a broad spectrum of generalized and unlabeled data and capable of performing a wide variety of general tasks such as understanding language, generating text and images, and conversing in natural language.

Large language models

Large language models (LLMs) are AI systems designed to process and analyze vast amounts of natural language data and then use that information to generate human-like responses. An LLM is any statistical model of language built on large data volumes and billions of parameters. LLMs are the basis for most of the existing foundation models, which increasingly are multimodal generating multiple outputs (such as text and images).

Parameters

A parameter is a configuration variable that is learned during a machine-learning process. Such variables are used to control the behavior of the model. LLMs have tens or hundreds of billions of parameters (GPT-4 is said to have 1.76 trillion). The number of parameters tends to be a measure of the size and the complexity of the model. The more parameters a model has, the more data it can process, learn from and generate. There is now a trend towards smaller, more specialized LLMs with fewer parameters and which are easier and cheaper to train.

Tokens and tokenization

A token is a basic unit of text or code that an LLM uses to process and generate language. Tokens can be characters, words or other segments of text or code. Tokenization is the process of splitting the input and output texts into smaller units that can be processed by the LLM AI models. Tokenization affects the amount of data and the number of calculations that the model needs to process. The more tokens that the model has to deal with, the more memory and computational resources the model consumes.

Training

AI training is the process of teaching an AI system to perceive, interpret and learn from data. Data scientists can spend years creating a new AI model and training it to perform complex tasks. In the first phase of a training process the model is fed massive amounts of data and is then asked to make decisions based on the information. The training part involves making adjustments to the model until it produces satisfactory results.

Fine-tuning

Fine-tuning an LLM involves making small adjustments to a pre-trained model to improve its performance in a specific task. This can yield better results than training a model from scratch as the model already works well and can leverage its existing knowledge to learn new tasks more quickly.

Prompt engineering

This is the process of refining LLMs with specific prompts and recommended outputs, and of refining input to various GenAI services to generate text or images.

Hallucinations

These are inaccurate or inappropriate responses to inputs presented as facts, such as false or misleading information or race/gender bias. AI hallucinations are often caused by limitations or biases in training data and algorithms, and can result in producing content that is wrong or even harmful.

Retrieval Augmented Generation (RAG)

RAG provides a way to optimize the output of an LLM with targeted information without modifying the underlying model itself. The targeted information can be specific to a particular organization and industry, so the GenAI system can provide more contextually appropriate answers to prompts.

(sources: AWS, Ada Lovelace Institute, IBM)

section 2:

who's who in the GenAI value chain?

When it comes to building commercial models a GenAI value chain is still emerging. That value chain – and the partnerships it is based on – is likely to change over time as new companies build their own LLMs and as successful GenAI applications developers seek to extend their roles to improve their own products and generate new sources of revenue.

For the time being, the GenAI value chain is dominated by hyperscalers. They are developing multiple LLMs, some generalist and some specialist, and embedding GenAI-enabled services in their enterprise platforms. Perhaps most importantly, they have the deepest pockets when it comes to investing in building and training LLMs and the computing power that supports them – as well as investing in GenAI start-ups and fledgling companies.

Infrastructure

The machine learning algorithms used to create text, images and audio through GenAI require vast amounts of data and fast memory to run effectively. This GenAI infrastructure space is dominated by semiconductor giants such as Nvidia, Intel and AMD, as well as the hyperscalers, some of which are developing their own chips (among others, Google, Amazon, Microsoft, Meta, Apple and Baidu are all developing chips).

The early-stage GenAI value chain

This graphic is representative of the early-stage GenAI market and is not intended as an exhaustive list of the companies operating in these areas

TM Forum, 2023

According to data from Omdia's Competitive Landscape Tracker, the semiconductor industry recorded a revenue increase in the second quarter of 2023 after five straight quarters of decline. Omdia says this growth is mainly attributed to the "explosion in the AI segment, led by GenAI".

"The data processing segment, driven by AI chips into the server space, grew 15% quarter-over-quarter and makes up nearly one-third of semiconductor revenue (31% in 2Q23)," Omdia notes. The research firm says quarterly revenue grew 3.8% to US$124.3 billion in the same period.

Cloud platforms

Hyperscalers develop the platforms that provide access to computer hardware that drives GenAI services. The three largest platform providers are Amazon Web Services (AWS), Microsoft Azure and Google Cloud. But other platform companies which can be expected to play a leading role in the telecoms industry's adoption of GenAI include Salesforce, Oracle, IBM and ServiceNow.

The sheer scale of investment needed to rapidly build and scale GenAI products makes it extremely difficult for other companies to compete. That said, a huge amount of private equity has flowed into start-up GenAI businesses.

According to The Wall Street Journal, investment in GenAI start-ups globally totalled $3.9 billion in 2022, rising sharply to $17.8 billion in the first nine months of this year ($10 billion of that by Microsoft into OpenAI, $270 million into Cohere's AI platform, and $100 million by SK Telecom into Anthropic). One of the leading developers of LLMs for GenAI, AI21 Labs, has received $336 million in funding from companies including Intel, Comcast, Google and Nvidia.

The immediate opportunity for hyperscalers to monetize their investments in GenAI is in enriching their existing products and services. For example, they could use copilot services to bolster sales of existing productivity tools. This would command an additional per-user fee. By opening copilots to developers, hyperscalers will also drive greater usage of their core cloud computing services.

Foundation models

This is the most dynamic part of the GenAI value chain. While hyperscale service providers are extremely active in building LLMs,

> **"Investment in GenAI start-ups globally totalled $17.8 billion in the first nine months of this year.**

many other companies – from start-ups to mature technology organizations – are seeking to build their own. Some of the most prominent companies here include:

- OpenAI – the company that owns ChatGPT and in which Microsoft has invested an estimated $13 billion.
- Anthropic, set up in 2021 by former OpenAI employees and which has developed Claude 2, a rival chatbot to ChatGPT. Amazon is reportedly investing up to $4 billion in the company and Google recently agreed to invest $2 billion.
- Stability AI, a company set up in 2020 which describes itself as "the world's leading open-source generative AI company". It has created the Stable Diffusion text-to-image model and Stable LLMs and has a partnership with AWS to run AI models on its Sagemaker machine learning platform.
- Cohere, an enterprise-focused GenAI firm set up by ex-Google Brain employees to deliver "language AI" (LLM) applications including chatbots, search engines and copywriting. It has key partnerships with Amazon, Google Cloud, Oracle and Salesforce.

New LLM initiatives are announced on a weekly basis, and the sector is trending towards more specialist LLMs which are cheaper to create / train and geared towards specific market segments and geographies.

Machine learning models

Generative AI models can be complex and expensive to train because of the computing power and resources needed. This makes it difficult to scale them to production environments. Machine learning operations (MLOps) helps to streamline the process of taking machine learning models to production, and then maintaining and monitoring them. It can help to scale AI models by automating the training and deployment process, and makes use of a range of tools to curate, host, fine-tune and manage foundation models.

Applications and services

We are only at the very early stages of products being developed for consumer and B2B markets that use GenAI either off-the-shelf or with a degree of fine-tuning. While the barriers to entry for companies seeking to enter the GenAI infrastructure, cloud platform or foundation model parts of the value chain are significant – because of the sheer level of investment required – the GenAI applications space is wide open.

Specialist knowledge in GenAI is already a valuable asset, and professional services firms are developing a range of services and solutions designed to help enterprises leverage the GenAI opportunity, including systems integration, training and professional services.

Telecoms and GenAI

So where does the telecoms industry fit into the GenAI value chain? What is clear is that both

telecoms operators and their technology partners see huge opportunities.

Analyst companies are also bullish about monetization prospects. McKinsey, for example, has identified 63 generative AI use cases spanning 16 business functions that it says could deliver total value in the range of $2.6 trillion to $4.4 trillion in economic benefits annually when applied across industries.

CSPs are already partnering with cloud platform providers – principally hyperscale service providers – and professional services firms as they take their first steps in GenAI. They are experimenting with off-the-shelf models with an expectation that they will train or fine-tune them for specific use cases. However, if operators do build their own LLMs – something which we explore in section 4 – they will need access to computer hardware in either a public or private cloud environment and build internal expertise in a range of MLOps tools.

CSPs' technology suppliers are also seeking to embrace GenAI tools and capabilities. Given that operators' own internal data – for example, data derived from customer interactions or from network operations – sits within the systems supplied by their operational and business support system (OSS/BSS) vendors, they will have a crucial role to play in how operators use GenAI.

The approaches of three of the companies sponsoring this research report provide good examples of how vendors will use GenAI:

- Netcracker is seeking to insert itself between the CSP and the LLM by ensuring that the LLM only uses the most relevant data – and that which adheres to rules governing security and privacy – when, for example, it receives a prompt from a chatbot.
- Tecnotree greatly increased its AI expertise with the acquisition of AI engineering platform CognitiveScale in December 2022. It is now embedding Gen-AI capabilities into its existing BSS stack. One of these is a chatbot that will enable its customers to ask questions about the Finnish company's products and solutions.
- Billing provider Aria Systems is partnering with Salesforce to offer a new AI-optimized "concept-to-care" monetization solution. This will allow CSPs to enhance products in their catalogs, integrate them into their OSS/BSS systems and offer automation and GenAI capabilities from Salesforce Einstein – the company's AI for CRM products portfolio – to personalize customer touchpoints.

In the meantime, operators are still working out how they fit into the GenAI value chain. The box and video below outline some of the early GenAI use cases being experimented with at Vodafone and Microsoft. And on the next page we outline TM Forum's work in AI and automation, including the development of a GenAI LLM model for telecoms.

In the next section we look at how some CSPs are experimenting with LLMs.

Vodafone and Microsoft experiment with early GenAI use cases

At DTW23-Ignite Vodafone's Head of New Technologies and Innovation, Lester Thomas, talked to Ian Thornhill, Telco Industry Advisor at Microsoft, about early use cases of GenAI and how TM Forum's Open Digital Architecture enables the operator to scale services using Open APIs. Thomas says Vodafone has its first GenAI use cases in production and "experiments in virtually every part of" its business, where any form of written or spoken human language is used, including supply chain, legal, customer operations and marketing. All of its AI is running in the public cloud, enabling it to focus on use cases and investment in skills.

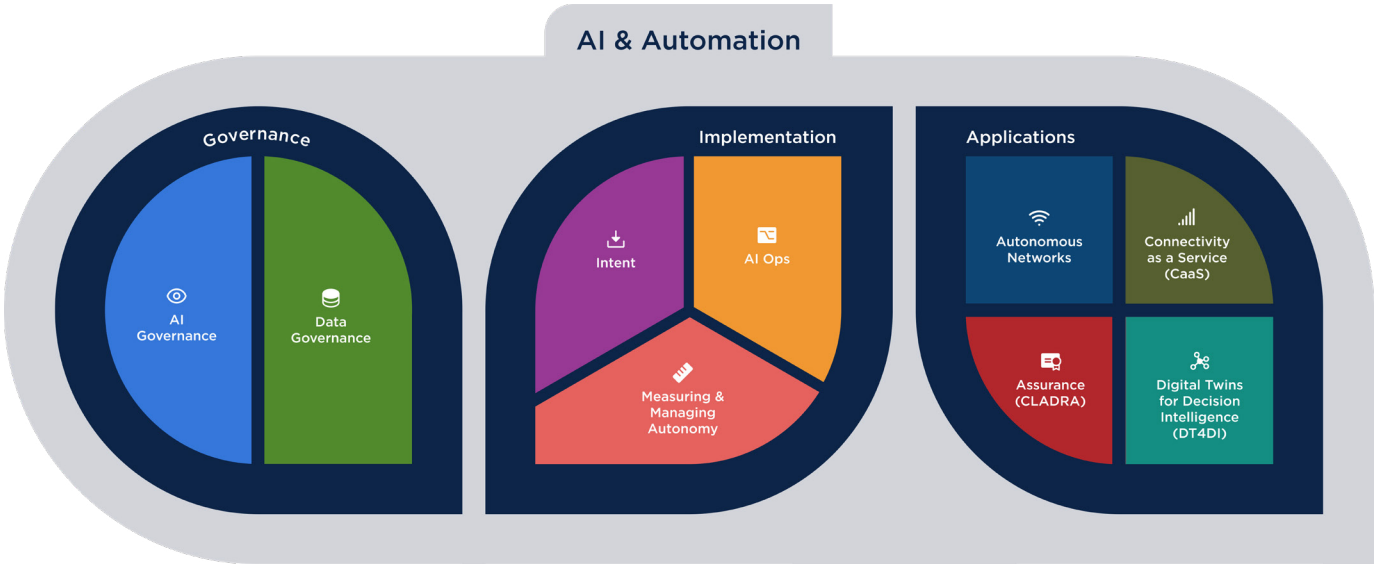
"Our goal is to enable every single employee across Vodafone, in every single function, to derive value, make faster decisions, by using insights and by using AI," says Thomas.

Thomas says in a few years he could see "virtually every piece of software, any part of your Open Digital Architecture", using LLMs.

Microsoft, among its applications for GenAI, is using its Azure OpenAI large language model as a tool to improve the linguistic capabilities of its chatbot and provide a "more natural language flow", says Thornhill. It is also using GenAI to capture and summarize data across different channels such as chatbots and call center agents.

Watch the video to find out more:

TM Forum's collaborative work in AI and automation



TM Forum, 2023

TM Forum members are working together in collaboration projects and Catalyst proofs of concept to solve common AI challenges for the telecoms industry. Central to that is a new Innovation Hub, hosted by Jio in Mumbai, which opened on 1 December 2023. Founding members are Accenture, Deutsche Telekom, Google Cloud, Orange, Reliance Jio, Telenor and Vodafone.

Among the early projects will be the development of TMF Guru, a GenAI/LLM conversational search capability to simplify and improve how to find relevant material digitally. The project members are setting out to create an architecture blueprint that outlines how GenAI technology can be seamlessly integrated into the telco landscape, addressing challenges related to security, privacy, accuracy, performance and scalability.

In the near term, TM Forum's AI innovation agenda focuses on gaining maximum benefits from AI, safely and cost-effectively, using existing IT and network systems and data. The medium-term goal is to exploit the full potential of AI to reduce operational costs and create opportunities for growth by taking AI from a bolt-on automation aid to an integral part of IT and network system designs.

The Forum's AI collaboration projects are creating standards and exploring best practice for AI-driven, intent-based closed loop automation, building an industry-agreed framework for the use of AI in IT and network operations. Meanwhile, Catalyst co-innovation projects are turning ideas into proofs of concept as members explore applications of machine learning and GenAI.

Members can keep control of AI deployments and avoid accumulating technical debt by following TM Forum's best practice in AIOps and AI governance, and prepare for the future by upgrading IT and network systems for AI with TM Forum's blueprint reference architectures, ontology and Open APIs. Many operators are adopting the Autonomous Networks Levels Evaluation Methodology to set targets for network automation and track their progress.

tmforum knowledge

REPORT

section 3:

CSP strategies – getting ready for GenAI

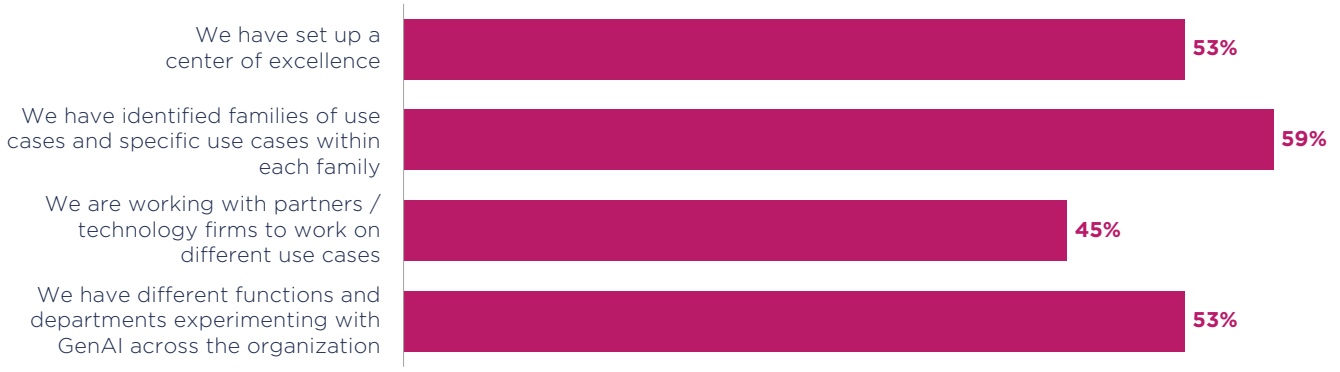
The arrival of GenAI tools has lit a fire under CSPs' AI strategies. But will its potential impact on operators' businesses be so significant that it requires them to put in place a new set of initiatives, a distinct and separate organizational approach and new investment?

Our survey respondents were split 50:50 in terms of whether GenAI needs a specific focus or if it should be seen as one strand within an overall AI strategy. The main argument for taking a specific focus is that GenAI tools can be deployed relatively quickly and provide near-term benefits – particularly in areas such as customer experience, but also potentially in many other functions across the business as we saw in section 1. As such, the value of this technology can only be unlocked by democratizing access and, potentially, budget. This may mean circumventing traditional approaches for putting together business cases and approving investment. It could also be argued – although this is somewhat controversial – that GenAI has a more transformative capability than other branches of AI and that, as such, it needs to be prioritized.

The counter argument is that the main use cases for GenAI are largely the same as those for other branches of AI and that creating parallel structures and initiatives

> **"The value of GenAI can only be unlocked by democratizing access and, potentially, budget.**

How are you approaching the GenAI opportunity (choose all that apply)?



TM Forum, 2023

would involve a duplication of effort and, potentially, unnecessary confusion and complexity further down the line. Furthermore, GenAI is often combined with other AI technologies and, as such, does not represent something that can be pursued as a standalone strategy.

OpenAI introduced the ChatGPT API for developers to integrate functionality in their applications in March 2023. That was the cue for CSPs to start putting together new use cases or, more likely, look at how it could help improve the results and performance of existing ones.

To develop these use cases many operators have created AI centers of excellence – 53% of our respondents say they have done so – in order to decide how, where and when to make the first moves. One of our respondents, as a first step, set up six work streams (six "Ps") managed by the AI center of excellence: potential platform, partners, people, process, and policy and guidelines.

Others are experimenting with GenAI across different functions and departments, and/or working with partners including technology firms to work on use cases. In the first instance many CSPs have reached out either to hyperscalers – many have established a "preferred" hyperscaler relationship within the past two-to-three years – or large consulting and professional services firms. This has given them a degree of comfort as they consider some of the key risk factors such as breaching rules around privacy, security or intellectual property; these firms have already considered the various issues as they build up their own capabilities.

In our survey, operators' own centers of excellence came out as the most important source of knowledge on GenAI followed by public cloud companies (see chart on the next page). Respondents were asked to rank six sources according to importance. Hyperscalers are already making GenAI services available alongside their existing public cloud services, making it easier for CSPs to dip their toes into the water. Industry analysts and consultants were the next most important sources.

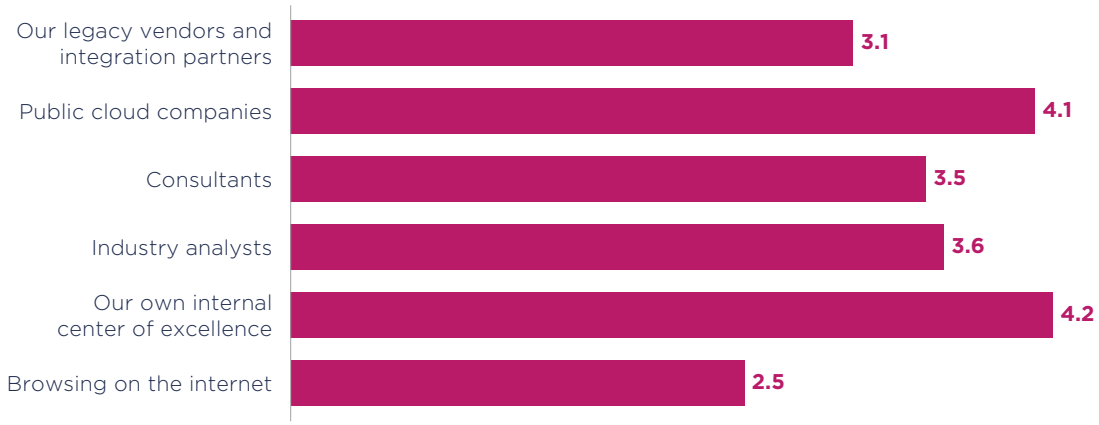
Speed versus risks

The launch of ChatGPT, Google's Bard and the explosion of interest in GenAI in the spring of 2023 has put huge pressure on telecoms operator CIOs to deploy the technology and demonstrate a return on investment in early use cases. Unlike other technologies, GenAI has captured the imagination of CEOs – and main boards – who are prepared to make resources available but are impatient for success.

Faced with the challenge of acting quickly, the CIO of one CSP in China – a mid-sized operator with around five million customers – decided to skip the process of developing use cases centrally and, instead, to start making GenAI agents available to any team that wanted to experiment with them. In the first month 200 teams said they wanted to try out the GenAI agents, and from this process the operator identified nine "flagship" use cases – but did not disclose them – where a business case could be made for driving revenues.

Another operator talked of the "risk and fear" that can be associated with deploying new technologies and how to overcome these to maximize the potential of successful experimentation: "We are trying to inspire our employees to start taking risks in a controlled way, and to do the experimentation and eventually record the outcome of these decisions so we can see the results of these use cases, and which have worked and which haven't worked."

Which are the most important sources of knowledge about GenAI for your business (rank most to least important)?



TM Forum, 2023

Getting a working proof of concept up and running as soon as possible was also seen as a priority by respondents to our survey. One western European CSP said that one of its early priorities, before receiving guidance and funding from the business, was to set up a working chatbot proof of concept. "Nothing speaks more than a working demo which actually shows you a bit of the magic," the company's Head of Data and AI told us.

The time that it takes to set up a working example can vary, however. This particular operator aimed to have working proofs of concept set up in just two days, while another operator we spoke to in our research said they had taken the whole summer to set one up.

Unlocking data value

Operators have spent many years trying to improve the availability, reliability and usability of data that their operations and customers generate. The commercial viability and availability of AI has made this a bigger priority for CSPs in recent years, and in our survey most operator respondents were relatively positive about the advances they have made in a number of different aspects relating to data readiness: the creation of a unified data model, data governance, the awareness of the data that is available across the organization, and the availability of data in real time.

> **"Nothing speaks more than a working demo which actually shows you a bit of the magic."**
>
> **Head of Data and AI, European CSP**

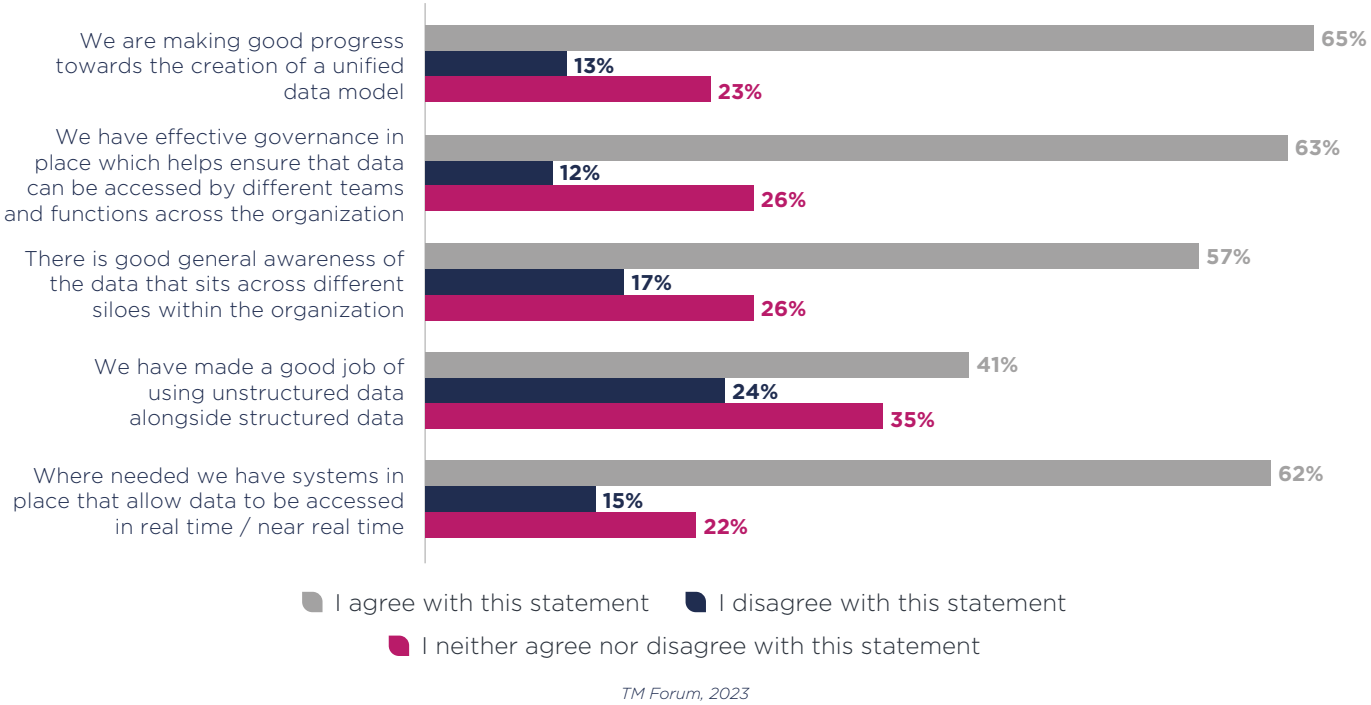
But when it comes to making use of unstructured data – for example, data generated from online or real conversations between customer care agents and customers – fewer respondents (41%) said that their organizations have made good use of the opportunity (see chart on the next page). This may represent an immediate opportunity for GenAI. Without GenAI the operator would need to label this data – the process of adding meaningful and informative labels to provide context so that a machine learning model can learn from it – to be able to use it. But now, with a large language model, an operator can feed this unstructured data into the model and immediately start generating insights which, over time, can be fine-tuned.

Getting ready for GenAI

In our research many senior executives stressed the importance of "organizational readiness" to pursue GenAI transformation. There are many elements to this, including:

- It is sometimes said that GenAI has the capability to democratize AI because tools such as copilots give individuals and small teams the opportunity to use new tools. But if access to data, both within and across siloes, is not easily accessible, democratization risks leading to frustration.

How successful has your organization been in making data available for AI and machine learning?



TM Forum, 2023

- The democratization of GenAI could also put the organization at greater risk from a range of issues associated with the technology, including hallucinations – incorrect or inaccurate responses to questions or tasks – fraud, security and broader legal considerations. Operators are already putting GenAI guardrails into place – for example, barring the exposure of raw GenAI outputs directly to customers without a human filter.
- Many of the business leaders we spoke to made reference to lessons that can be learned from the early days of cloud transformation where many operators took a "lift and shift" approach. These operators then found that the cost of running applications and workloads in the public cloud did not generate the benefits they expected and that public cloud costs spiralled way beyond what they budgeted (more on this in the next section).
- Such is the speed of development of GenAI that operators are faced with the challenge of how to keep their organizations moving at such a pace to allow them to be flexible and agile in terms of how, where and when to adopt new solutions and technologies.

In the AWS-commissioned Altman Solon survey, 19% of respondents said they had use cases in production, with this number set to increase to 48% in the next two years. We look in more detail at GenAI use cases in section 5, while in the next section we examine how CSPs are experimenting with LLMs.

> **"The democratization of GenAI could put an organization at greater risk from a range of issues associated with the technology.**

tmforum knowledge

REPORT

section 4:

To build? To train? To fine-tune? Getting to grips with LLMs

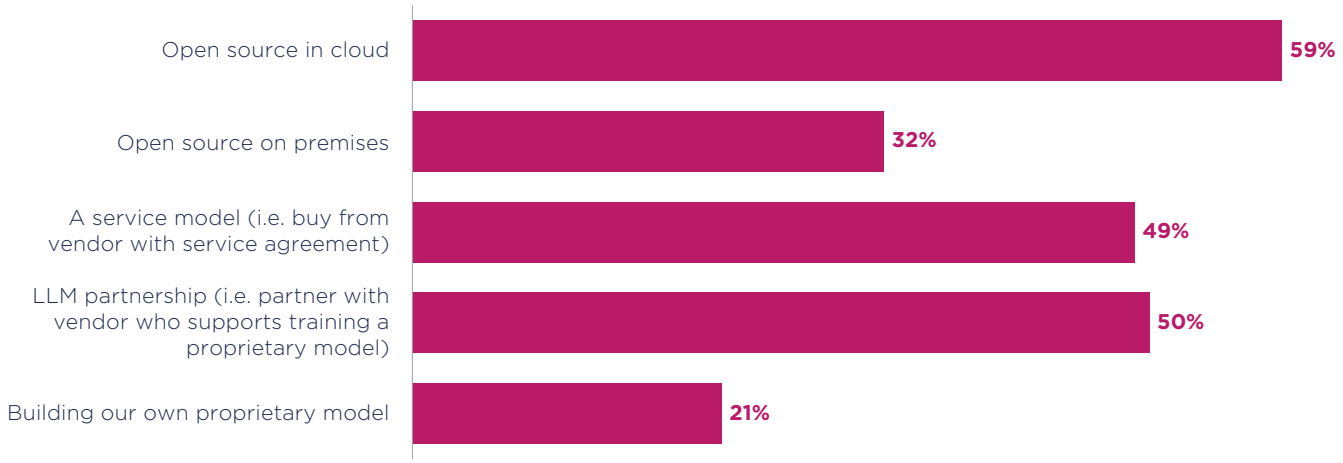
CSPs face a number of choices when it comes to choosing foundational models, the use cases to which they are applied, the extent to which they need to train or fine-tune them, and the environments in which the LLMs operate.

There are two categories of generative AI LLMs: proprietary and open source. Proprietary models are expensive to develop – various estimates put the cost of GPT-4 at $50 million to $100 million – and anyone who uses them needs to get a license. Open-source LLMs are freely available for people to access, use and tweak based on their own particular requirements.

Examples of open-source models are Llama 2, developed by Meta and Microsoft; and GPT-Neo and GPT-J from EleutherAI, a collective of researchers working on open-source AI; Examples of proprietary models are BloombergGPT, developed for the financial sector, and DeepMind's Gopher.

The size of an LLM is a function of the data on which it is trained and the number of parameters. An LLM that is trained on general internet data is necessarily much larger than an LLM that is trained on domain-specific data. A domain could be a specific industry vertical – for example, healthcare or financial services – or a horizontal segment such as research papers from academia. Some LLMs contain a combination of internet data and domain-specific data.

Which large language model (LLM) approaches are you considering for your organization (choose all that apply)?



TM Forum, 2023

Minimizing investment

In this early, initial phase, many CSPs want to use LLMs that require minimal up-front investment and which can generate immediate benefits. That means choosing the right use case mapped with the right LLM and from a supplier that can provide support and guidance. In such instances, where external skills are needed, open-source LLMs might not be a viable option.

CSPs also need to decide whether the LLMs they use should operate on their premises, on a public cloud or on vendors' clouds. The decision-making process here is similar to any other compute requirement. An on-premises solution can be attractive when dealing with sensitive data, but public cloud is a better option if the operator is experimenting or needs to scale computing resources up or down quickly. In practice, given the dominance of the hyperscalers, the vast majority of LLM usage by CSPs will likely be in the public cloud, at least in the short-to-medium term.

In our survey we asked CSPs which LLMs they were considering for their organizations. When it comes to using open-source models, 59% said they were considering the use of that option in the cloud while 32% were looking at on-premises options (see chart above). About half of respondents said they were considering a service model – an approach where the CSP buys access to an LLM from a vendor along with a service agreement. A similar number are considering partnerships with LLM vendors which can support them in training proprietary models.

> **"CSPs need to decide whether the LLMs they use should operate on their premises, on a public cloud or on vendors' clouds.**

In the AWS / Altman Solom survey more than 65% of respondents said they anticipated training off-the-shelf models, primarily using their proprietary internal data to tailor them to their specific needs. Only 15% said they would build foundation models in-house.

There were no clear trends in terms of preferences for CSPs using context learning / prompt engineering (i.e inference) or fine-tuning (i.e. training) to train models. Most CSPs are evaluating both approaches. However, in this early phase operators seem to be leaning towards context learning as this is less resource intensive and provides "good enough" performance for many use cases.

Choosing between LLMs

While it is true to say that one foundation model can be used for multiple tasks, it is also a reality that one LLM is not going to do all the tasks that operators are considering. But how do CSPs make decisions between the tens or even hundreds that are available to them?

The only way for operators to inform their decisions is to experiment with different models, gauge the accuracy, the bias and the level of hallucinations – inaccurate or inappropriate responses to inputs presented as facts by the model – that come with each of them. CSPs also need to learn about the training data that was used for each of the different models and how this impacts their accuracy. As new, domain-specific models are launched they may decide that these yield better results than more general-purpose ones, even though the latter likely have benefitted from greater training investment.

In deciding how good or how accurate an LLM is, a CSP will consider whether, for example, it is 40%, 60% or 80% accurate. Different use cases will require different levels of accuracy for the use of GenAI to be worthwhile. In some cases a CSP may decide that, for example, 20% accuracy is worthwhile if it creates insights that can be input into a process or decision. For others, 70% or 80% accuracy may be needed.

The processes of training, fine-tuning and Retrieval Augmented Generation (RAG) – optimizing the output of an LLM with targeted information – are used to improve the accuracy of LLMs. But telcos also need to learn how LLMs work – and how to involve different teams and departments – in order to get the most out of them. The more parameters in an LLM the more training will be required and, as such, the more expensive it will be to develop because of the required computing resources.

> **Recruiting new skills and training staff scored the highest among operators wanting to exploit AI.**

"We're trying to figure out how we enable our domain people to understand the nuances of a generative model," the CIO of one CSP in the Indian subcontinent told us. "Each model has its own weights and its own methodology."

As such, when it comes to fine-tuning and prompt engineering this is not solely an area that sits within the remit of a data science team. It also needs to involve domain specialists. When it comes to fine-tuning the model, this process involves understanding how the model works and then framing questions in such a way as to generate the most accurate results.

In our survey we asked what operators need to do to best exploit AI. Recruiting new skills, and training staff in AI skills, scored the highest among our respondents, while partnering with AI specialists was another popular option (see chart on the next page).

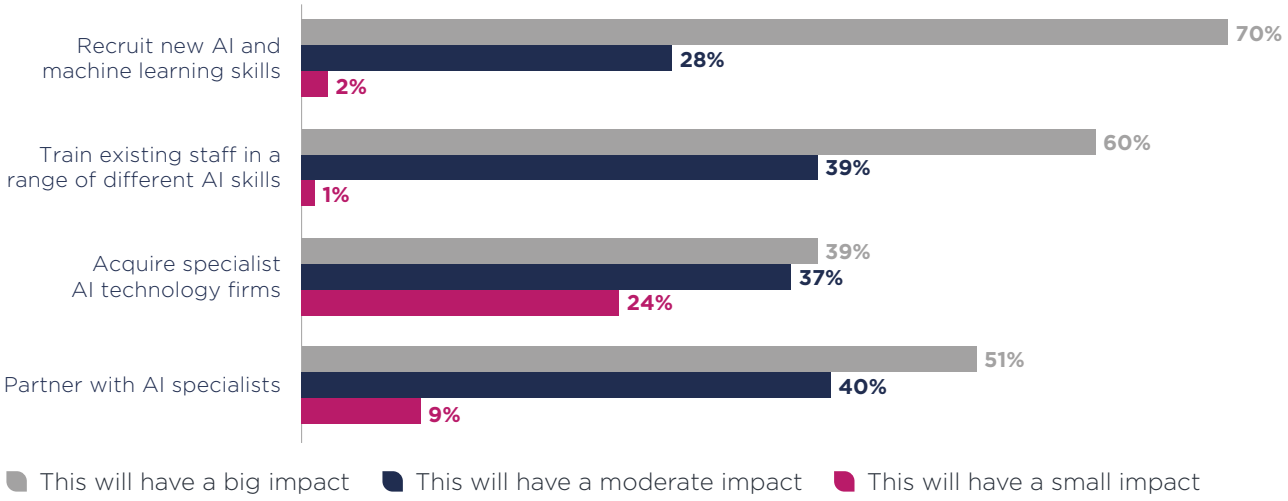
CSPs are also finding that LLMs behave differently from each other and that each one has its own "personality". This means that, for example, when comparing the performance / accuracy of one LLM versus another, initial results can be disappointing until it becomes clear how best to use it.

Choosing the right large language model(s)

Cost & licensing	Is the cost model fully understood, particularly if / when use cases scale and are there any restrictions on usage?
Performance of the LLM	Has the model been assessed in terms of its accuracy, usefulness, human evaluation and fidelity?
Use case requirements	Can the LLM fulfill the specific tasks / use cases that are required?
LLM capacity	What are the computational and memory requirements of the LLM?
Pre-training data and domain	Has the model been pre-trained on general internet text or domain-specific data?
Fine-tuning flexibility	Can the model be adapted to a specific task using an organization's own-labelled data?
Ethical and responsible AI considerations	Is the model committed to fairness, bias mitigation and responsible AI practices?
Updates & maintenance	How frequently is the model updated and how is the model maintained by the model operator?

TM Forum, 2023

What do operators need to do to give themselves the best chance of effectively exploiting AI and machine learning?



TM Forum, 2023

Other considerations when choosing an LLM include:

- **Language availability.** Some models are trained on a single language while others are trained on many. GPT-3, for example, is trained in more than 50 languages. In smaller countries, or ones where there are many different languages, CSPs may need to take some of the responsibility themselves for training different aspects of LLMs.
- **Cultural nuances** expressed in languages. If we take, for example, the language used between a customer and a customer services agent, there are very different types of language and conventions used across the world.
- **The longevity of the model.** LLMs have a limited life span because there is necessarily a cut-off point in terms of the data that is fed into a model. The evidence to date is that a model only has a lifespan of around six months. This is a particular issue for fast-moving industries where the output of the model needs to reflect data that was recently fed into it. Furthermore, there is no guarantee that when an LLM model developer releases a new version it will behave in the same way.
- **Level of latency supported.** Do the use cases that the LLMs support require low-latency connectivity? If so, this may impact the choice of private versus public cloud.
- **Ease of transferral / cross-use.** How easy will it be for a CSP that has trained a specific model to transfer its input into a new or updated model?

Cost and ROI

CSPs in general are comfortable experimenting with LLMs and meeting the payment terms of their owners. While companies that embed LLM capabilities tend to charge for them on a per-user, per-month basis – Microsoft's copilot service is made available for an additional fee to its Teams customers – businesses seeking direct access to LLMs pay on an API basis. Calculations are, essentially, based on the number of words input into or output from a model. As such, costs scale directly with usage.

Such pricing models present a significant cost management risk to any enterprise that deploys LLM-enabled solutions at scale. This is a particular issue with GenAI, which is seen as a tool to "democratize" AI – and innovation more broadly – across the organization.

The discussion around cost and return on investment (ROI) has echoes of early deployments of public cloud services. Many operators took a "lift-and-shift" approach to moving applications to the public cloud and subsequently found that the business case was no longer viable because the workload consumed more public cloud resources than expected.

As such, operators are keen to evolve to a different type of pricing model. An ideal model – from the perspective of operators – would be one that incentivizes

the LLM model owner based on a successful deployment of an LLM for a specific use case or family of use cases. That could be, for example, the number of customer queries successfully handled by a chatbot or a reduction in the time taken by field engineers to address network outages.

Watch Verizon CDIO, Shankar Arumugavelu, and TM Forum CEO, Nik Willetts, discuss the comparison of GenAI adoption with the journey to cloud-native operations.

Telecoms-specific LLMs

Financial information services provider Bloomberg unveiled BloombergGPT, an LLM for the financial sector, in April 2023. Bloomberg has taken a "mixed approach" by combining financial data it owns – largely the data sources accessible by any user of a Bloomberg terminal – with data that it has scraped from the internet. This latter category comprises both financial data and more general content. The LLM has 50 billion parameters constructed with 363 billion tokens from Bloomberg's own information sources, and 347 billion from other, mainly publicly available, datasets.

If Bloomberg is able to create an LLM specifically for the financial services industry, does this mean that it should be possible to create one for the telecoms sector? SK Telecom – in partnership with Deutsche Telekom, Singtel and e& – is now building an LLM for the telecoms industry (see box on p.23). But is this the right medium-to-long-term approach for the industry, or indeed the only approach?

In our survey just 21% of CSPs indicated an inclination to build foundation models in-house. But why shouldn't operators have their own model? The graphic below highlights some of the arguments for and against.

In the next section we look at some of the early use cases for GenAI in the telecoms sector.

Arguments for CSPs having their own model

- With IP ownership, a CSP may be able to license the model to other CSPs, or simply to make it part of the overall technology capabilities of the business.
- By creating its own LLMs, the telecoms industry could uncover new products and services leading to revenue growth.
- A dedicated telecoms-industry LLM will, almost certainly, produce significantly more accurate results because it has had real-world data fed into it. This is particularly true for data that is specific to telecoms – for example, network operations. The ability to use GenAI to generate code could produce significant benefits.
- With their own LLM, CSPs can have control over their own destiny rather than relying on third parties which do not necessarily share their own best interests.
- Each CSP use case has its own requirement in terms of latency, privacy and security – major considerations when it comes to deciding whether to deploy on-premises, or on private or public cloud. With their own LLMs, CSPs can make the best choices.
- Pricing models for LLMs deployed at scale are unclear. "The industry is pretty much in its infancy in terms of deciding how they charge for the use of LLMs at scale," the CIO of one North American CSP told us. By developing their own LLMs, operators will have a better visibility of costs and ROI.

Arguments against CSPs having their own model

- Within an operator's business many different functions may benefit from GenAI adoption. But each has a distinct approach to language: customer-facing functions use highly conversational language; technology teams deal in technology terms; finance teams have their own financial language, as does legal, and so on. Would each of these functions be better served by a domain-specific LLM?
- The telecoms industry may not need to build its own LLM model to obtain (telecoms) domain specificity. This may be possible though the process of training or fine-tuning existing models.
- Operators are slow-moving organizations which are not used to developing and deploying their own technologies at anything close to the speed at which the AI industry is evolving. There is a risk that the challenge to innovate, to change direction and to build in new features and capabilities will mean that an operator-specific LLM will fail to keep up with more general models.
- The cost to build and train their own model tailored to the telecoms sector, or specific segments of the sector, could be a cost burden. Not having clear visibility of costs or ROI could be prohibitive.
- CSPs may lack the in-house skills to build and train telecoms-specific LLMs.

Alliance sets out to build telco-specific LLM

South Korean operator SK Telecom has been developing its AI capabilities for the past two-to-three years using its own super computer, Titan. In August 2023 it invested $100 million in GenAI start-up Anthropic (also backed by Google). As part of the deal, the companies agreed to jointly develop an LLM that is optimized for telecoms operators.

This LLM, which is due for launch in early 2024, is a variant of Anthropic's more general model called Claude and will support Korean, English, German, Japanese, Arabic and Spanish languages. The model will be shared by three other operators – Deutsche Telekom, e& and Singtel – which have formed, with SK Telecom, the Global Telco AI Alliance.

Both Deutsche Telekom and SK Telecom emphasize the early focus of the new LLM will be on helping them deliver a stronger customer experience. An LLM's reasoning capabilities make it good at analyzing customer intent and searching a telco's back-end systems for the answer, says Jonathan Abrahamson, Chief Product & Digital Officer at Deutsche Telekom.

SK Telecom Chief AI Global Officer Chung Suk-geun told attendees at TM Forum's DTW23-Ignite event in Copenhagen in September that with GenAI CSPs had the potential to redefine the relationship with their customers. In doing so, CSPs not only have the opportunity to build more loyal relationships, but they can also leverage this to sell new products and services. One example would be to use the skills used in improving customer experience through GenAI to help enterprises outside the telco industry, he said.

Watch the video to find out more:

Michael Lesniak, Business Development and Partnerships Manager for Web 3 at SK Telecom, reckons that GenAI could, more specifically, help operators develop their carrier billing – the sale of third-party services – line of business. "I think that GenAI creates an interesting opportunity for carriers to use SMS to communicate directly with customers and to initiate transactions," he told attendees during a webinar organized by Mobile Europe.

Rather than making the customer go through the sometimes long-winded process of signing up for a new app on a smartphone, an operator could remove all the friction from the transaction because it knows enough about the identity of the customer to open up access to the service without going through the usual process, because they already have a billing relationship with them.

The four operators will build a so-called Telco AI Platform that sits on top of the new LLM and which will serve as the foundation for developing new services such as chatbots and other apps.

Speaking at SK Telecom's financial results in August, Suk-geun said part of the thinking behind the creation of a telco-industry LLM was that operators are having problems getting the attention of the companies building LLMs. The Global Telco AI Alliance plans to add new operator members to help drive scale and increase the richness of the model and the applications that are built on top.

Giving operators more control of their business, and properly leveraging new technologies, is also part of the thinking behind building their own LLM and AI platform. By working together operators will be able to avoid the mistakes that they have made in the past in ceding strategic business to rivals during a major technology shift, according to Suk-geun. This point was echoed by Deutsche Telekom's Abrahamson, who said the company "liked the idea" of developing and owning its own IP.

Suk-geun says operators will start by using GenAI to transform internal core processes such as marketing, sales and customer service operations. For example, he says, marketing cost reductions could come through automated customer service and in networks through automated infrastructure monitoring; and service quality improvements could be made through more personalized, targeted offerings.

tmforum knowledge

REPORT

section 5:

taking the first steps – early GenAI use cases for CSPs

The key question many CSPs are asking is what practical applications can be applied to telecoms operations using GenAI technology. But before exploring the potential use cases, and families of uses, it is worth considering how it is currently being tested by early adopters.

Early GenAI deployments will mainly be used for improving existing functions. The technology cannot, currently, generate content that is sufficiently or consistently accurate or suitable to replace complex operations. However, as developments improve it will start to be used in this way, and in so doing will replace processes, functions and people.

However, we are still some way away from this point. Indeed, it is likely that either businesses themselves, or policymakers, will impose strict rules and guidelines to restrict GenAI use.

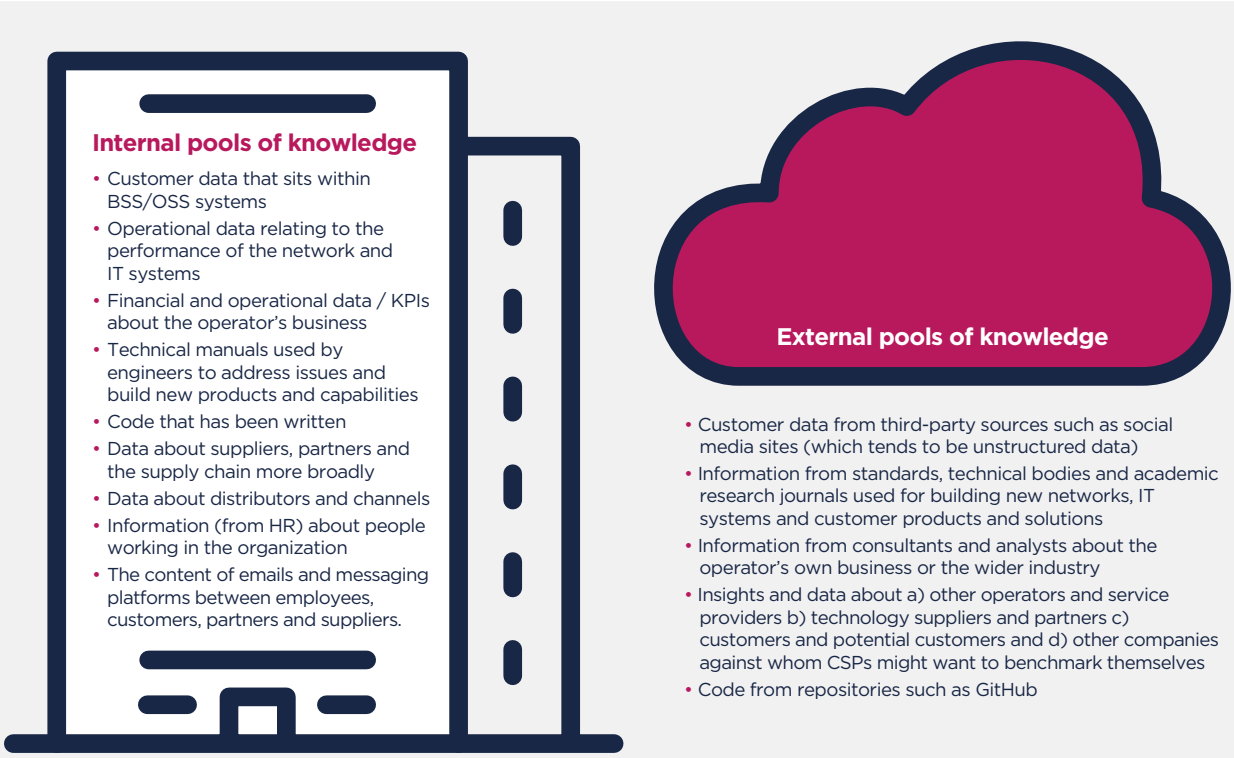
With early use cases mainly taking advantage of out-of-the-box LLMs, operators are able to deliver improvements to existing functions with minimal investment. CSPs that we interviewed for this research are confident that these can generate a rapid, easily demonstrable return on investment.

The cost savings for GenAI use cases that involve the replacement of existing functions can be significant, but they can also generate other benefits – for example, delivering customer service capabilities that CSPs do not currently offer because they are too costly. But for these to become viable the accuracy and fidelity of LLMs needs to improve significantly.

For GenAI to deliver value it needs a body of knowledge to work on. With telecoms, pools of knowledge exist both inside the operator organization and outside. The graphic on the next page defines some of those aspects.

Once CSPs become familiar with GenAI's capabilities they have the opportunity to create new pools of data as a clear and deliberate strategy to deliver future customer service, operational efficiency and productivity gains, and to deliver new or enhanced products.

> **"The cost savings for GenAI use cases that involve the replacement of existing functions can be significant.**



Families of use cases

Many operators and their technology partners have begun the process of putting together families of use cases for GenAI. We have identified seven for this research:

- customer operations
- sales and marketing
- network (operations)
- IT and software engineering
- product innovation
- internal knowledge
- business operations.

Within each of these families we have identified different use cases which are either being explored by CSPs today, or have short-to-mid-term potential. For many we include graphics from the Altman Solon survey. It asked CSP respondents to say whether they were implementing or evaluating GenAI across 16 use cases in four of those categories: customer operations, sales and marketing, network operations, and IT and software engineering.

Customer operations

Delivering a better chatbot experience for customers is seen as the most tangible opportunity today for CSPs investing in GenAI. Many operators already offer their customers access to chatbots as a complement to other forms of customer interaction. But these services are extremely limited in the functionality they provide and are not, generally, liked or appreciated by customers.

With GenAI, operators have the opportunity to completely change the chatbot experience. Today's services are based on traditional AI and rules-based systems. As such it is useful for certain queries and frequently asked questions. Indeed, these chatbots can be effective whenever the system recognises a query from a customer and the data that sits behind it can point the customer to information that helps to resolve the problem.

GenAI chatbots would use unstructured rather than structured data to understand what the customer wants and how it can help. These multi-modal chatbots would open up use cases such as predictive customer sentiment.

Today's LLMs, however, are not accurate enough to give CSPs sufficient confidence to deploy these types of chatbots. Operators would need to train large volumes of customer data, either on a third-party LLM or their own, to give it the capability to have meaningful conversations with customers.

In the shorter term, and with less accurate information, customer service agents can save themselves time by using GenAI copilots to search for information to help with customer queries.

Building families of GenAI use cases

Customer operations
- Customer chatbot
- Call center agent documentation and coaching
- Website assistance
- Predictive and personalized services

Sales & marketing
- Marketing collateral generation
- Personalized customer / email scripts
- Social media automated responses

Network
- Field service operations guided assistance
- Network / capacity planning
- Network security testing
- Post Mortem Creation
- Root cause analysis

IT / software engineering
- Automated code generation and testing
- Automating repetitive tasks (e.g. data mapping)
- Detection of code security vulnerability

Product innovation
- Carrier billing
- Personalized services
- Voice value-added services
- B2B customer call services

Internal knowledge, training & development
- Evaluating new trends / developments
- Competitive analysis
- Supply chain analysis

Business operations
- Legal / contract
- Fraud management
- Partner management (e.g. roaming)
- Human resources

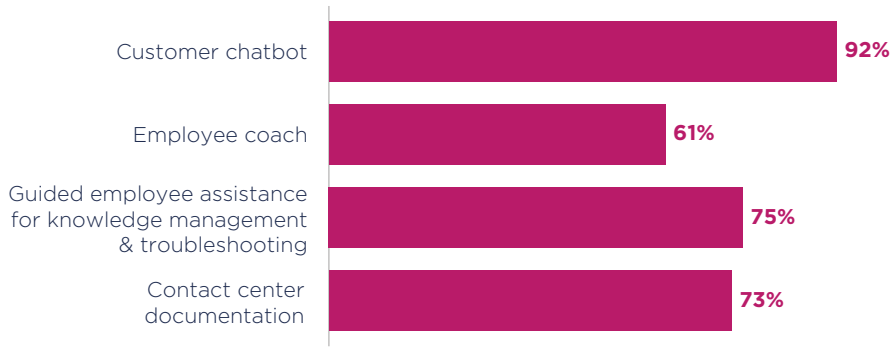
This can help to bridge the performance gaps between those agents who are able to find the information that a customer needs and those who are not.

The copilot approach could also be something that is offered to customers – for example, when they are searching for information on a website. One North American operator who took part in our survey has combined three technologies – conversational AI, enterprise search and GenAI – to create a multi-modal experience for its customers. The role of GenAI, in this example, is to provide a summary of what the customer is looking for (and finds) on the website.

Canadian telco Telus, meanwhile, is using GenAI to predict when customers will have negative experiences – for example, when the network is likely to go down because an engineer is carrying out repairs in the local area.

Matthew Sanchez, Global Chief Data and AI Officer at BSS vendor Tecnotree, says his company is currently tracking around 20 use cases in customer operations. "Some of these we're working on actively with customers, some are discussions we are having with customers, and others are R&D work we are doing. We have seen some RFPs around that involve using GenAI to fully transform customer operations."

CSP GenAI use case adoption: customer service



TM Forum, 2023 (source: Altman Solon)

The Altman Solon survey identified four use cases within the customer service category. Its responses, which measure the percentage of total respondents that are currently implementing, or evaluating with a high likelihood to adopt, the use case demonstrate that chatbots are, by some margin, the application which is attracting most interest.

Sales and marketing

The highest profile use case within the sales and marketing category is automated generation of marketing collateral / content. It allows B2B marketers to build compelling narratives, tailor messaging to target audiences and, ultimately, drive engagement. It represents an important productivity tool as marketers come under growing pressure to create personalized messages for target audiences. Within this overall category of content marketing, GenAI can help marketers to brainstorm and generate ideas, create new content, enhance existing content and create compelling visuals.

As customers start to use their own preferred social media

channels to communicate with their service providers, GenAI can help CSPs to respond to these queries. However, because of the very nature of social media this is fraught with risks if a response is, in any way, inappropriate or incorrect. A CSP would need to have an extremely high level of trust in its LLM to allow direct GenAI-created content to communicate with a customer.

When it comes to sales, GenAI gives account managers the ability to send personalized email to customers and prospects. It can also help sales teams identify the best prospects and create scripts for conversations with them.

Network (operations)

Whereas many of the use cases in customer operations and sales and marketing are generic to all businesses, the network is specific to telecoms. As such, use cases will emerge from operators themselves, often in partnership with GenAI technology firms.

Networks are a major focus for the deployment of AI-enabled automation more broadly, and GenAI will be used both to complement existing focus areas and to address new opportunities.

Near-to-mid-term opportunities include what Altman Solon describes as "guided employee assistance for installation, troubleshooting and maintenance". Deploying field service operations teams to install network elements such as routers and switches to maintain the network and to respond to outages is an extremely costly function. But the time it takes for network engineers to resolve issues can be shortened if they are given access to tools that enable them to quickly diagnose issues by consulting manuals and historical data gleaned from similar issues that have occurred in the past. This capturing of historical events, which can then be used to learn from, is also referred to as post mortem creation.

Another use case within network operations is capacity planning. Smart capex is an analytics and AI-driven approach already used by many operators to make more efficient decisions about when and where to invest in new network capacity. GenAI can be used to enhance and fine-tune smart capex by adding unstructured data from customers – from conversations with customer service agents, comments on social media and so on – to more structured network data.

There are many other potential applications for GenAI within the larger network planning category – for example, network design. But CSPs are already using other types of AI for these issues, and it is not always clear how much GenAI can add. Furthermore, such is the sensitivity of network data that operators' default position tends to be to proceed with caution.

Root cause analysis is an approach already used in CSPs' technology functions – and in enterprise IT more broadly – to solve issues by going to the root cause of faults or problems. By using GenAI to interrogate either internal or vendor documentation, CSPs have the possibility of automating the processes that are used to evaluate and classify alarms. This can help ensure that operators prioritize and address critical issues and generate recommendations for their resolution.

CSP GenAI use case adoption: marketing and product

TM Forum, 2023 (source: Altman Solon)

CSP GenAI use case adoption: network (operations)

TM Forum, 2023 (source: Altman Solon)

An LLM can generate network topology diagrams or build a digital twin based on text descriptions, and learn patterns from historical data to identify anomalies and the root cause of problems. It can also automate troubleshooting and predictive maintenance. China Telecom, for example, is building its own GenAI system to find the root cause of network problems, using structured and unstructured data.

The Altman Solon survey also includes network image generation – a capability that assists with the guided employee assistance use case – and synthetic data generation for security testing. The creation of synthetic data came up as a strong theme in many of our conversations and is a capability which will help CSPs to implement many different use cases across their organizations.

Synthetic data is information that is generated artificially rather than produced by real-world events. Typically created using algorithms, it can be deployed to validate mathematical models and increasingly is used to train AI / machine learning models. Synthetic data is also often used for product testing and software development.

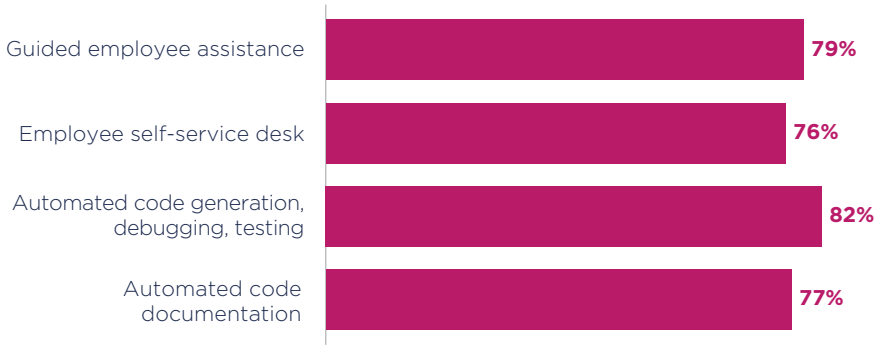
Software engineering

This is an area where the use of copilots, or companions, can help software engineers and developers to be more productive. For example, AWS's AI coding companion uses a foundation model to generate code suggestions in real time. Its suggested use cases for CSPs include automating software development with text/voice to code, filling skills gaps and helping people without coding experience.

Sanchez at Tecnotree says a significant opportunity for CSPs lies in automating many of the repetitive tasks that sit within CSP operations. "Data mapping, data migration, these are constant problems in the industry," he says. "People transferring data from one place to another, validation, logic, a lot of business rules and business assumptions that have to then get coded into rules and code. Generative AI can help with a lot of this."

AT&T has given its employees access to a tool called Ask AT&T, initially built using ChatGPT. The first use case is "helping our coders and software developers across the company become more productive", says Andy Markus, AT&T's Chief Data Officer, in a blog. Another use case it is exploring is upgrading legacy software code and environments.

CSP GenAI use case adoption: software engineering



TM Forum, 2023 (source: Altman Solon)

> **"A significant opportunity for CSPs lies in automating many of the repetitive tasks that sit within operations."**
>
> **Matthew Sanchez, Tecnotree**

Product development

Much of the interest in using GenAI for product innovation is based on the potential for CSPs to build better relationships with their customers and to learn about their tastes and preferences. Speaking at the DTW23–Ignite conference in Copenhagen in September, SK Telecom's Chung Suk-geun gave the example of a telecoms operator offering a customer travelling abroad personalized restaurant recommendations.

Redefining customer relationships is one of the two overarching strategies that SK Telecom is pursuing with GenAI (the other is internal transformation). It has a vision for rolling out GenAI enabled chatbots that learn about their customers' tastes and preferences and, in doing so, offer them personalized services.

Voice services also represent an opportunity for CSPs to add more "richness" and potentially to build new revenue streams, particularly in B2B. Voice calls could be transcribed and translated in real time and then stored and archived. Roaming presents a specific opportunity, allowing

customers to converse with local people using their own spoken language. This builds on a concept called 5G New Calling, developed by China Mobile.

Internal knowledge

Every function that sits within a telecoms operator business has a requirement for knowledge, training and development. Today this is delivered via company intranets, third-party literature and documentation, and online and offline training and development programs. GenAI can help drive the adoption and effectiveness of these initiatives with, for example, the adoption of employee chatbots and LLMs that deliver assistance in areas such as:

- Competitor analysis
- Staying up to date with new trends in technology, marketing and so on
- Partner and supply chain background / conformance assessment.

The main challenge here is in updating the information to train the model. Competitive analysis can only be truly useful – unless it is used for general background purposes – if it is updated on a daily basis.

Business operations

In addition to the six families of use cases we have listed so far, many other general uses are emerging from across the CSP business. A few examples are:

- GenAI can be employed to create synthetic datasets that mimic real-world fraud patterns. These synthetic datasets can be used to augment limited or imbalanced training data, improving the performance of fraud detection models. TM Forum published a whitepaper on this topic in September, and Canadian operator Telus is already using GenAI for fraud detection.
- Vodafone is building an intelligent assistant that uses GenAI to securely and quickly search through the 10,000+ contracts that it has signed with other telecoms operators for mobile roaming services.
- The legal departments of CSPs can use GenAI to help prepare contracts by feeding previous ones into an LLM.
- HR departments could use the technology to help classify and rank job candidates by feeding their CVs into a GenAI model.

Below and on the next page we look at some of the TM Forum Catalyst proofs of concept experimenting with GenAI use cases. And in the next section we look at some of the risk factors and potential regulation surrounding the deployment of GenAI technologies.

Using GenAI to deliver hyper-personalized customer experience

GenAI Hyper-personalized Customer Experience was a finalist in the **Outstanding Catalyst – Business Impact** category in the TM Forum 2023 Catalyst Awards. The catalyst focuses on the future of customer experience using hyper-personalized real-time GenAI.

The aim is to address the problem of duplicate and unnecessary messages to customers, thereby reducing churn. The solution is to send personalized messages through a real-time marketing automation pipeline that combines predictive and generative AI. It is predicated on three steps:

- carrying out complex event processing
- using predictive AI to predict a churn score and then instantly create a dynamic campaign syndicated through an omnichannel experience
- using generative AI to complete digital engagement.

The participants say the design pattern could be used for other applications such as network healing, agent workflows, bi-flow automation and in the overall ethical use of data in AI technology evolution.

There were five CSP champions: Bell Canada, BT, Deutsche Telekom, Telecom Italia and TurkNet. The solution, based on TM Forum's Open Digital Architecture and Open APIs, used Google Cloud's Vertex AI platform, Pegasystems' Customer Decision Hub, and Accenture's Customer Data platform.

Watch the video:

Assessing Scope 3 emissions to deliver lower-carbon product offers

The Digital Carbon Footprint Optimization catalyst was the winner of the **Best Moonshot Catalyst – The Energy Challenge** at DTW23 – Ignite. The project sets out to tackle the problem of measuring Scope 3 emissions – indirect emissions in the supplier value chain – by helping CSPs determine value chain emissions data and easily transpose that data to a product catalog. The project uses real-time data and AI (including generative AI) built on sustainable cloud infrastructure to create a new way to build product bundles and provide ongoing carbon dioxide equivalent (CO2e) data and analytics.

The solution enables customers to have full visibility of the carbon footprint of products such as phones during purchase. A built-in intelligent tool analyzes CO2e data using TM Forum Open APIs in various aspects of the shopping lifecycle, enabling the CSP to provide lower carbon footprint product offers and publish it across different digital channels.

The catalyst demonstrates how to monitor consumers' battery consumption and battery health and provide recommendations based on AI / machine learning. Consumers thereby get full transparency into the carbon emissions generated by their devices.

Using data from sustainability consultancy Carbon Footprint the project team was able to provide access to accurate, real-time carbon footprint information for each product component. CSP product managers and marketing teams were able to view CO2e data for each component, ultimately allowing them to create more environmentally oriented product bundles which make use of the CO2e visualized data attributes in the product catalog.

The project champions were Vodafone and Carbon Footprint, with technology partners AWS, Amdocs and Snowflake.

Watch the video:

GenAI helps deliver 5G intelligent operations

To achieve the level of network intelligence needed in the 5G era, CSPs increasingly are turning to AI to detect, diagnose and resolve network performance issues. The **GenAI and knowledge-driven 5G operations** Catalyst set out to develop an intelligent cognitive decision-making system for 5G network operation, with the aim of improving network performance, reducing operational costs and enhancing customer experience.

The project champions China Telecom Research Institute and China Unicom Research Institute, together with their technology partners, are developing a solution to implement 5G intelligent operations, using generative AI to construct LLMs and a communication knowledge graph to perceive the quality and condition of the network and its services. The system can classify and perform root-cause analysis of problems, and then execute automated resolutions.

The system will automatically collect and process network data, classify and analyze root causes through a pre-trained AI model, and automatically delimit and locate problems in combination with a network resource database. It then matches the solution through the decision knowledge base, forms the intention with the resource base, and sends it to the network management system to identify the intention and perform operations on the network.

Building on TM Forum's Open APIs, the solution enables intelligent closed-loop automation to create and sustain a network capable of advanced self-cognition. As well as advancing the intelligence and efficiency of network operations, the solution can enable a range of other savings such as reducing 5G network operations and maintenance (O&M) costs.

Technology partners for the Catalyst included ZTE, Inspur Group and Beijing ZZNode Technologies.

Watch the video:

tmforum knowledge

REPORT

section 6:

look while you leap – key challenges for CSPs

We have identified five categories of challenges relating to CSPs' use of GenAI. In some cases operators are facing these challenges today, while in others they are likely to be encountered in the medium term. All relate to the deployment of GenAI at scale.

The categories we highlight are broad and there are many more challenges within them. But they are among the key issues CSPs will need to address to make full use of GenAI technology in the telecoms space.

Access to (CSP) internal data

CSPs have long struggled to organize, store, aggregate and extract the data that sits in their organizations and within their business and operational support systems (BSS/OSS). This should be no surprise given the fact that those support systems have evolved over a period of 20-30 years and have been supplied by many different vendors.

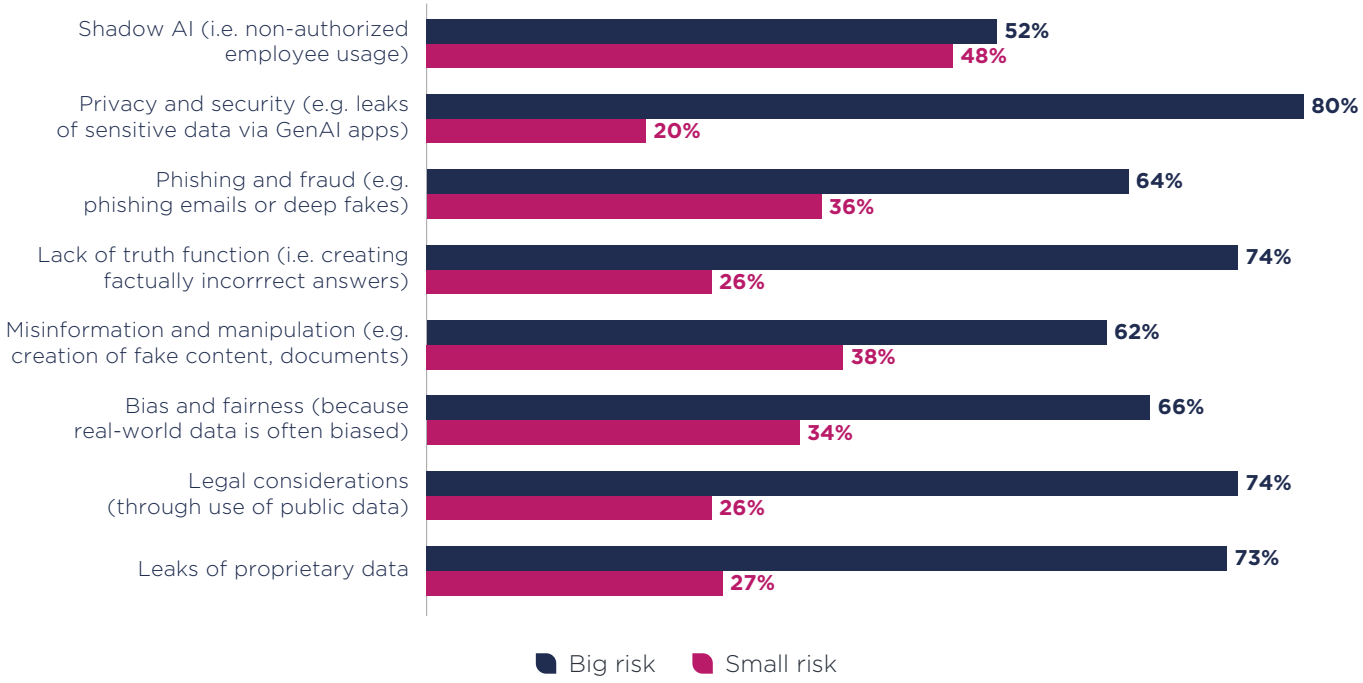
Nevertheless, considerable progress has been made in terms of moving towards a unified data model – a common structure for sharing data – making it available across siloes and delivering it in real time. Furthermore, these challenges to accessing the right data are applicable to all forms of AI, not specifically GenAI.

The suitability of unstructured data for GenAI is an important differentiator with other types of AI. CSPs have, until now, been relatively unsuccessful in using unstructured data – for example, messages, transcriptions of voice calls or articles on the internet – in AI implementations.

There are many early use cases in GenAI – particularly ones that use copilots – that exploit unstructured data. But most GenAI use cases will still need structured data from operators' support systems (Netcracker estimates that 90% of GenAI use cases need BSS/OSS data). The challenge for operators is how to integrate this sensitive customer data with the data that sits in public LLMs, and how to ensure its accuracy given that it is being drawn from both public sources – which cannot necessarily be trusted – and operators' own more trustworthy data.

Another data-related challenge facing CSPs is making data available to non-technical individuals and teams who lack the knowledge or training to consult internal systems which can often be complex and hard to navigate.

What are the main risks for your organization of using GenAI?



TM Forum, 2023

Security, governance and privacy

Data security and privacy was cited as the number one big risk in our CSP survey by 80% of respondents (see chart above). Given the strong focus on using GenAI for customer operations, CSPs need to be fully cognizant of where the servers running the relevant LLMs are located. There are stiff penalties in many countries for sending customer data – including anonymized customer data – outside of national borders. This means, for example, that because OpenAI runs its LLMs over Microsoft Azure, using customer data in GPT-3 or GPT-4 is only possible in those countries where Microsoft Azure has public cloud infrastructure.

Regardless of where the LLM is running, the approach of customizing pre-trained models and tailoring them with operators' (sometimes proprietary) data is one that requires a robust data infrastructure and governance to ensure that the foundation model is trained on high-quality data. Even when the right safeguards and governance are in place, the very nature of GenAI means that when it is deployed at scale, and across the organization, new systems and processes may be needed.

> **"If it's something you're going to scale you need to think about other things such as change management."**
>
> **Charlotta Lundell Berg, Telia**

"If it's something you're going to scale you need to think about other things such as change management," says Charlotta Lundell Berg, Head of Analytics at Telia.

And Ibrahim Gedeon, CTO at Telus, believes the sheer speed of innovation enabled by GenAI is putting existing security and privacy processes under pressure. "We used to get one, two or three projects [that require a process to evaluate the data] every year, but now with GenAI we're getting one or two every week," he says. This is forcing the operator to put more agile processes towards governance into place.

As governments and regulators across the world consider the potential impact of GenAI on their economies and societies, many will introduce new safeguards in an effort to reduce some of the bigger risks. The European Union, for example, has proposed a regulatory framework for AI which

specifically references GenAI. Disclosing the content that is generated by AI, designing models to prevent them from generating illegal content, and publishing summaries of copyrighted data use for training, will all become mandatory requirements for systems that use GenAI under the proposed AI Act.

Phishing and fraud is another risk that can be put into this category and was cited by two out of three of our survey respondents as a big risk. Similarly, legal considerations such as the use of public data which may be implicated in legal or license questions is a major concern.

Other risks asked about in our survey sit within the category of data accuracy which we highlight below.

IP and lock-in

As we saw in section 4 there are two types of LLMs: proprietary ones owned by companies that have built them, and open source. As such, the IP of an LLM is owned by its creator. But what happens when a company uses an LLM under license and adds to it through the process of fine-tuning? The operator may own the data it has added, but this does not mean it is possible, or easy, to transfer that data or the learnings from the process of fine-tuning the model from one LLM to another. An operator may want to change which LLM it is using because the model is being updated or simply to use a different LLM.

Such considerations strengthen the argument for CSPs to develop their own LLMs. Indeed, a senior manager in a hyperscale service provider, who asked not to be named, told us he believed the telecoms industry "should own data and build the model". But as we saw in section 4, there are arguments for and against such approaches.

> **The use of public data which may be implicated in legal or license questions is a major concern.**

Issues over IP and the transferability of fine-tuning and learning activities also impact the approach that operators will take over how to customize an LLM for their particular requirements. Other approaches include retrieval augmented generation (RAG) – a way to optimize the output of an LLM with targeted information without modifying the underlying model itself – and prompt engineering (see page 9). The latter is a specialized AI skill that involves guiding and shaping model responses by carefully crafting specific and detailed questions, and testing out different ways to phrase instructions to generate better outputs.

Data accuracy and traceability

Many factors contribute to an LLM generating information which is unbiased and accurate. The key question is how accurate the data needs to be to meet the requirements of a particular use case. Less accurate data can still be valuable if the alternative is no data at all, or if it is being used to help guide someone in making a decision or taking a specific course of action. But much more accurate data is needed to replace human decision-making and interaction.

Accuracy is a function of the data that is fed into a model and of how the model uses and interprets it. Bias exists when the totality of the content on the internet on a particular topic does not show a suitable balance of purported facts or opinions. Hallucinations occur because the output of LLMs tends to be presented in such a convincing way that it is easy to accept it as fact.

Accuracy can also be undermined by a lack of up-to-date information – if, for example, there has been a cut-off date for information that has been fed into a model. The cut-off date for GPT-3, for example, was originally September 2021, but this has now been addressed through RAG.

"Reasoning" errors can be more difficult to address. LLMs apply statistical analysis to large bodies of text, but they are not logic engines. Furthermore, different LLMs can produce very different responses to the same interrogation or prompt. This can distort results when, for example, a CSP is making accuracy comparisons between two LLMs. The way that a prompt is phrased may create a misleading impression that one LLM is better than another when the reverse may be true if the wording is changed.

The ability to trace the output from an LLM is also extremely important, particularly if outputs are showing sub-standard results and the user needs to trace that back to the original source of information. Traceability can also be important for issues around privacy, security or legality.

Cost and ROI

The cost of experimenting with LLMs is not prohibitive and is not a factor impacting how CSPs are learning how, where and when to develop early use cases. However, it could well become a factor when it is deployed at scale.

Rather than extending the API-based pricing approach being used today, and which scales directly with usage, many CSPs would like to see a transition to new pricing models based on the success or outcome of a specific use case. However, hyperscale service providers, given the colossal investments they are making in GenAI, will come under enormous pressure to adopt pricing approaches which enable them to show investors that they are starting the process of making returns on their investments.

The visibility that a CSP has about costs varies depending on the use case. Michiel van Rijthoven, Lead Data Scientist at VodafoneZiggo in the Netherlands, cites the example of a customer chatbot using GenAI which, because there are clear metrics in place, enables the operator to gauge costs. "We know how many calls we get [into the call center], we know the length of the calls, so we can make some predictions about the cost. And yes, the cost will be significant, but the business case is extremely positive," he says.

In the longer term there is a risk that operators will incur technical debt in GenAI because of the growing array of choices that are emerging in how to generate better results from LLMs. As we have detailed, these include fine-tuning, RAG and prompt engineering. If the usage of these techniques is not properly monitored and managed, operators could find themselves in a situation in the future where it becomes difficult to adopt new, efficient techniques at scale because of the cost and difficulty of migrating old systems.

> **"Many CSPs would like to see a transition to new pricing models based on the success or outcome of a specific use case.**

In the final section we look at some of the key findings from our research for this report as well as some recommendations for CSPs setting out to experiment with and deploy GenAI LLMs.

section 7:

key findings and recommendations

Throughout this report we have focused on the opportunities and challenges to CSPs considering, and starting out on, the deployment of generative AI large language models. This section brings together those findings. The figures quoted in this section are based on our survey of 104 executives from 73 operators.

CSPs are responding in different ways to the GenAI opportunity. Here are some of the early trends:

- Despite the fact that GenAI – as a commercial proposition – is only a year old, **more than half of CSPs are experimenting** with it today.
- The mandate for launching into GenAI is, in many cases, coming from the main board. The desire to act "hard and fast" on GenAI is putting intense pressure on technology leaders. They now need to work out how to **align their existing AI strategies and focus with these new, bigger ambitions**. In our survey there is an even split between those operators which believe that GenAI requires a specific, distinct focus and those which see it just as another type of AI that can be incorporated into an existing strategy.
- **CSPs are focused today on low-hanging fruit** – use cases that can be exploited using the copilot services offered by, for example, hyperscale service providers and with out-of-the-box LLMs fine-tuned with their own data.
- Operators today are **experimenting with both proprietary and open-source LLMs**. The landscape is constantly shifting, and operators will experiment with many different models.
- **GenAI is ideally suited to working on unstructured data** and offers an immediate opportunity for CSPs. Many have struggled until now to make good use of the unstructured data that sits within, or adjacent to, their organizations.
- Today's API-based LLM pricing models allow CSPs to experiment, and to build proofs of concept without particular concern about the costs they are incurring. However, as and **when operators deploy GenAI at scale such pricing mechanisms may become less attractive**.
- Some use cases for CSPs involve the use of GenAI as the only type of AI, but many **require GenAI to be combined with other types of AI**. For example, operators are already deploying AIOps in areas such as network operations, and GenAI will be a complementary technology with the potential to deliver improved performance.

Main challenges

- There are many **internal issues that CSPs need to address** before they can deploy GenAI at scale. These include siloed data, the lack of a common ontology and incomplete information architectures.
- While concerns about LLM accuracy – for example bias and hallucinations – represent important challenges today, new techniques are emerging that will deliver significant improvements. As such, they may represent less of a long-term concern for CSP businesses than **issues around privacy, security, IP and the cost of deploying GenAI at scale**.
- **The ability to measure model performance is a significant challenge in GenAI**. In the world of predictive AI and machine learning, this was done through techniques such as precision and recall. However, approaches in AI are still being developed.
- There is some caution within the CSP community about rushing headlong into adopting GenAI for network use cases. In addition to sensitivity around networks data there is a view that a large body of work is already being undertaken by CSPs to drive AI-enabled automation in network operations, and that **GenAI is more of a complementary technology than a transformational one**.

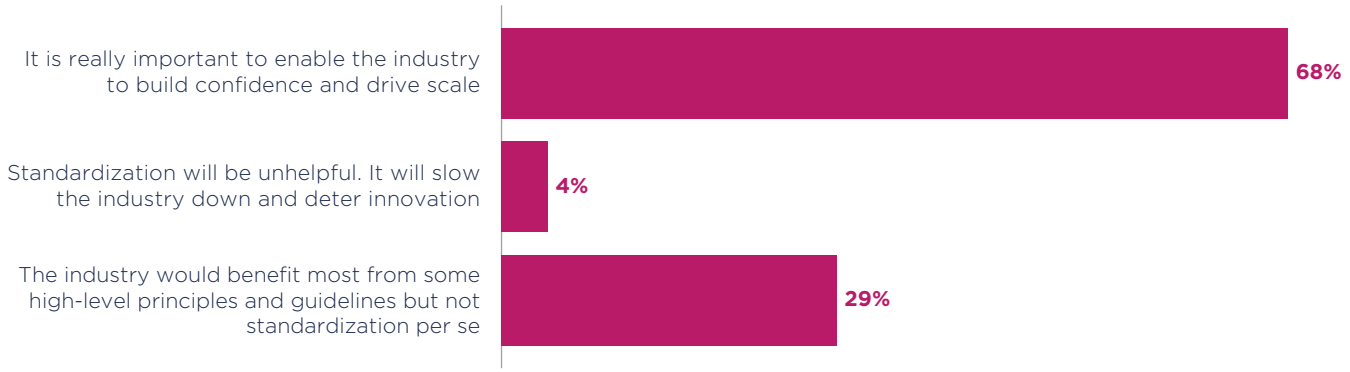
Main opportunities

- **Six out of ten of our respondents believe that GenAI will have a significant impact** on their business within the next 1-2 years and 37% within the next 2-5 years.
- In our research we have identified **seven families of use cases**: customer operations; sales and marketing; network (operations); IT and software engineering; product innovation; internal knowledge, training & development; and business operations.
- **Customer chatbots represents – by some margin – the use case that is being pursued most enthusiastically by CSPs.** The use of GenAI chatbots represents a completely different approach from current chatbots and a migration from a rules-based, structured data system to one that is based on unstructured data.
- The **relationship that CSPs have built with hyperscalers** in recent years puts them in a much better position to benefit from, and to help shape, GenAI-enabled product development than was the case, for example, in the early years of public cloud computing. Tier-2 and tier-3 operators, which do not have these relationships, are at a significant disadvantage.
- Rather than merely viewing LLMs as "databases" of information, **the key value lies in their reasoning ability** where the model actually decides for itself how and where to find information to solve a specific request or address a specific use case.
- The trend in LLMs is for **smaller, domain-specific models**. A domain could be a vertical sector – such as financial services – a country or a horizontal business function.
- **GenAI will not replace any of the AI technologies that have been deployed so far**; rather, it will complement them. As such, CSPs need to be able to combine the unstructured data that tends to sit outside of their BSS systems with the structured data that sits inside them.

Key recommendations

- In response to excitement within the business about the transformational potential of GenAI, and concerns within society about misuse of the technology, **operators need to adopt a mantra of aggressive experimentation** and thoughtful implementation.
- CSPs need to find the right balance between, on the one hand, building families of use cases and applying GenAI to them, and, on the other, democratizing access to GenAI to allow different teams to discover their own use cases. **GenAI should be seen as a tool for innovation within the wider CSP business**.
- Operators may (or may not) develop their own LLM(s). But it is highly unlikely that it/they would be used for the entirety of their business. **There may be more of a case for building LLMs in areas where there is telecoms specificity** – for example, the network. In more generic functions the speed of innovation in GenAI will likely outpace what CSPs can achieve with their own internal GenAI technologies and platforms.
- **CSPs should use the learnings and parallels from their public cloud journey to inform their approach to GenAI.** When CSPs first looked at the financials around public cloud migration the business case was extremely attractive. But when they shifted different workloads and applications to the public cloud the efficiencies did not always

How important will standardization be to make the most of AI and machine learning?



TM Forum, 2023

emerge. As and when operators deploy GenAI at scale they will **need to deploy tools to ensure that costs are managed efficiently**.

- **In order to scale GenAI adoption**, CSPs should make a distinction between the near-term benefits that GenAI can bring to their businesses, by using off-the-shelf LLMs and applications built on top of them, and medium or long-term issues that need to be addressed. These include cost and ROI, the risk of vendor lock-in, the potential for the creation of technical debt, and the integration of structured and unstructured data.
- As CSPs transform their networks and IT systems using cloud-native principles, and adopt architectural concepts such as TM Forum's Open Digital Architecture (ODA), **it is important that AI/GenAI becomes an integral part of the system design** rather than a "bolt-on".
- To truly leverage the potential of GenAI, **CSPs need a robust, detailed information architecture** which details what knowledge they have and how and where it can be accessed. They also need a new data architecture that specifies where data lives, how it is accessed by AI, and how frequently it is refreshed. TM Forum is developing a new data architecture as part of the evolution of the ODA. This will build on SID, TM Forum's existing Information Framework asset.
- The early promise of GenAI gives CSPs a glimpse into the opportunities for transforming customer relationships. Operators have struggled until now to deliver powerful digital experiences because a large proportion of customer interaction is still conducted via call centers and, to a lesser extent, retail stores. **GenAI offers the prospect of building better, closer, more lucrative relationships.** CSPs can fully exploit the potential of GenAI by using new customer chatbots, developing other messaging channels, and leveraging voice conversations to understand, diagnose and address customer queries and to upsell and cross-sell.

additional features & resources

tmforum knowledge

REPORT

Netcracker
An NEC Company

Bring Telco Domain Knowledge to GenAI with Security and Control

With its remarkable ability to create new content and engage in powerful human-like interactions, GenAI has the potential to dramatically advance all areas of the telecom business. It can bring a much needed step change in customer experience and satisfaction and also transform the business internally with significant operational and productivity improvements.

GenAI needs telco-specific knowledge to be effective

GenAI models (LLM/FMs) are only as good as the data they are trained on. The best LLMs on the market have excellent communication and content creation skills, but they have no knowledge about the complex world of telecom with its many domains, layers, technologies, platforms, business processes and highly sensitive data.

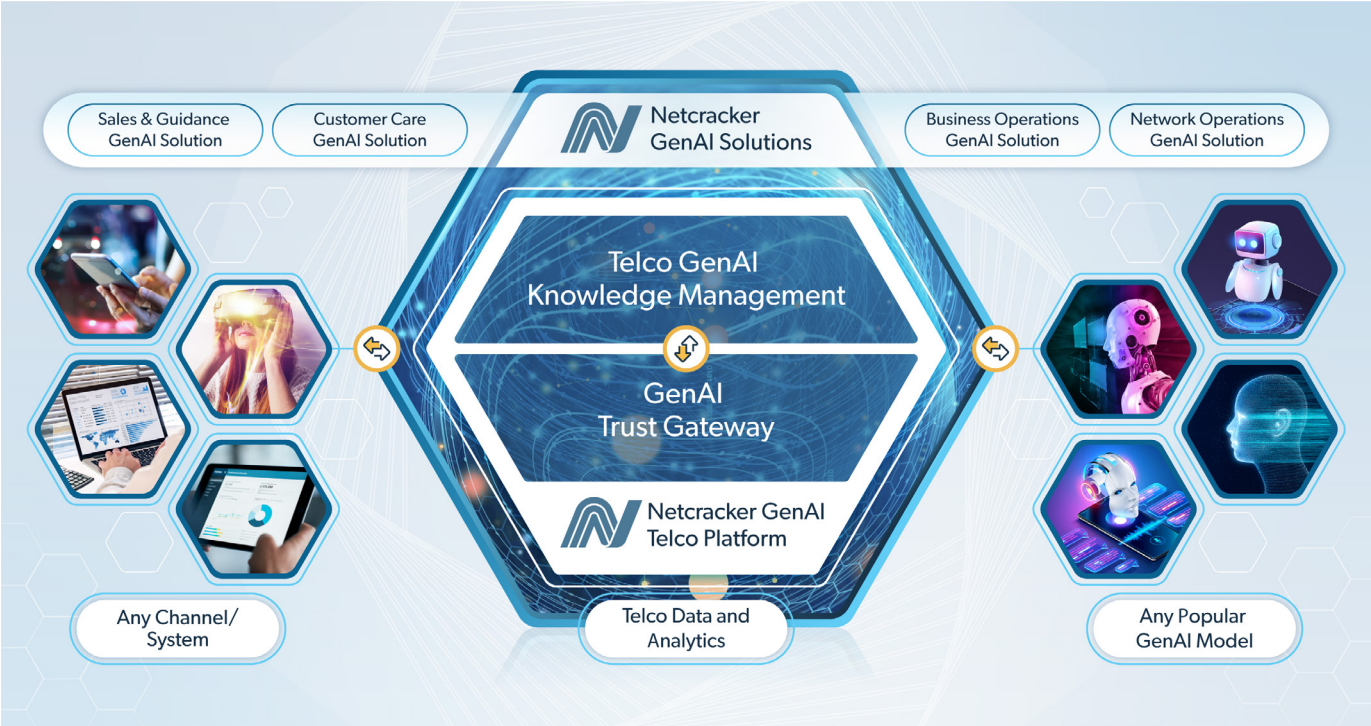
Approximately 90% of telco data comes from BSS/OSS and knowledge base systems including billing history, usage data and tariff plan information for digital assistants and agents, or catalog rules, offers and promotions for digital marketing assistants, and inventory data, assurance and configuration management for digital technicians.

Accessing this data in a secure and controlled way is critical to obtain any real value from this ground-breaking technology.

Bridging the telecom gap is challenging

Bridging the gap between GenAI models and the telecom business poses many challenges. It requires an integration with BSS/OSS and data analytics systems with the following challenges:

- The most advanced LLMs are public models, hosted in the public cloud, which poses an inherent security risk. They cannot be trained on sensitive telco data.
- A good proportion of telco data is real-time or constantly changing and therefore unsuitable for the fine tuning techniques employed for relatively static domain data.
- Creating custom GenAI models, with the quality of ChatGPT, is extremely costly in terms of training and running the model, thereby making it unaffordable for many CSPs today.



Netcracker's approach to unlock GenAI's value

Netcracker solves these issues with its GenAI Telco Solution, which sits at the intersection between GenAI models (LLMs and other FMs), the GenAI users (customers, partners, telecom employees and even systems) and the highly valuable proprietary telco BSS/OSS data and knowledge bases.

Netcracker's solution is governed by the following key principles:

- GenAI models need to be enriched with real-time telecom data and context to maximize the quality of the response and protect proprietary telco data.
- No personal or sensitive data can be leaked to GenAI public models to ensure the strictest security.
- CSPs will need a mix of GenAI models (public and private) that are suited to different tasks and price points.

Grounding GenAI models with real-time telecom data and knowledge

To overcome the issues related to real-time data and lack of telco business knowledge, a different approach is needed that augments each user prompt with additional data, context and instructions to increase the quality of the response.

Netcracker achieves this using techniques such as prompt engineering and retrieval augmented generation (RAG) to access its extensive telecom knowledge base systems and call up real-time BSS/OSS data, ingesting this additional information into the prompt. This approach, together with fine tuning the model, grounds the LLM/FM with relevant domain data making the responses highly accurate.

Establishing the highest levels of security and control

Security is cited as one of the major barriers for take-up of GenAI in telecom. The telecom business has very strict confidentiality rules due to its vast customer data as well as regional laws including GDPR. CSPs are particularly concerned when using the public models, or hosting private models in the public cloud.

Positioned between the GenAI models and telco data, Netcracker's platform prioritizes customer privacy by implementing state-of-the-art anonymization techniques. All sensitive customer and proprietary telco data is obfuscated, eliminating leaks to public models.

In addition, Netcracker's platform maintains tight control over API access to BSS/OSS data, providing a secure and regulated gateway that protects sensitive information while facilitating seamless data integration.

Maximizing ROI with a mix of GenAI models

GenAI can be applied to many tasks across the business – whether it's responding to a user question in a chat, or creating a new customer facing design, or generating a unique image for a marketing campaign. There are many models available on the market, each focused on a specialty. By choosing the right mix of models best suited to the task and its level of complexity, CSPs can maximize ROI.

Netcracker has adopted a unified approach for any GenAI model. Our platform is pre-integrated with many popular GenAI model aggregators including Microsoft (with Open AI), AWS Bedrock, and Google Cloud's Vertex AI as well as model specialists such as Cohere, Hugging Face, Anthropic, Midjourney, AI21 and Stability.ai.

Getting started with high-value use cases

Netcracker has collaborated with its customers and partners to create a set of ready-made use cases that solve the most demanding problems. The use cases include a range of digital assistants – both internally and customer facing – and work in conjunction with predictive AI and automation technologies to advance all areas of the telco business. Examples use cases include:

- **Digital Care Assistants** explaining billing details, new offers and promotions, and solving network issues
- **Sales Advisers** guiding B2B or B2C customers with new offers and personalized recommendations to close opportunities faster
- **Business Advisers** aiding with catalog configurations, business analysis and marketing content
- **Digital Operations Technicians** assisting with fixing network issues, configurations and integrations.
- **Service Design Assistants** creating new customer-facing designs.

Netcracker GenAI Telco Solution: Bringing telecom domain expertise to GenAI

Netcracker combines its telecom IT domain expertise and BSS/OSS market leadership with breakthrough advances in AI technology to make GenAI highly valuable and safe for the telco business.

With our deep understanding of the telecom industry, pioneering GenAI platform, and native integration with BSS/OSS systems we can help CSPs to deploy powerful use cases that will:

- Lower call center costs while increasing customer satisfaction, resulting in faster time to resolution and lower cost per contact.
- Increase internal business productivity for rapid completion of tasks, higher levels of innovation and greater self-sufficiency.
- Fill automation gaps with automated service designs, configurations and network/services fixes.

GenAI's potential is game changing but only if operators incorporate it into their processes in a secure and safe manner, resulting in improved productivity, lower costs and a competitive advantage in today's telecom market.

About Netcracker

Delivering the Digital Future
Netcracker
An NEC Company

Netcracker Technology, a wholly-owned subsidiary of NEC Corporation, has the expertise, culture and resources to help service providers around the world transform their businesses to thrive in a digital economy. Our innovative solutions – including our flagship cloud-native Netcracker Digital Platform – value-driven services and unbroken delivery track record of three decades help service providers to achieve their digital transformation goals, drive the telco to techco evolution within their organizations and realize business growth and profitability. For more information, visit www.netcracker.com.

tmforum knowledge

REPORT

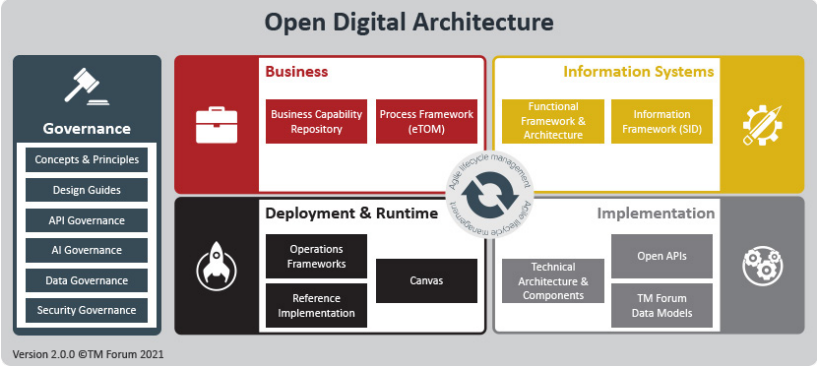
tm forum open digital framework

A blueprint for intelligent operations fit for the 5G era

The TM Forum Open Digital Framework (ODF) provides a migration path from legacy IT systems and processes to modular, cloud native software orchestrated using AI.

The framework comprises tools, code, knowledge and standards (machine-readable assets, not just documents). It is delivering business value for TM Forum members today, accelerating concept-to-cash, eliminating IT & network costs, and enhancing digital customer experience.

Developed by TM Forum member organizations through our Collaboration Community and Catalyst proofs of concept, building on TM Forum's established standards, the Open Digital Framework is being used by leading service providers and software companies worldwide.



The framework comprises TM Forum's Open Digital Architecture (ODA), together with tools, models and data that guide the transformation to ODA from legacy IT systems and operations.

Open Digital Architecture

- Architecture framework, common language and design principles
- Open APIs exposing business services
- Standardized software components
- Reference implementation and test environment

Transformation Tools

- Guides to navigate digital transformation
- Tools to support the migration from legacy architecture to ODA

Maturity Tools & Data

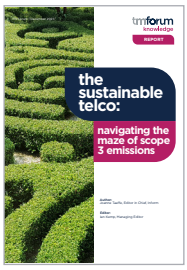
- Maturity models and readiness checks to baseline digital capabilities
- Data for benchmarking progress and training AI

Goals of the Open Digital Framework

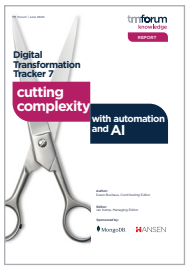
The aim is to transform business agility (accelerating concept-to-cash from 18 months to 18 days), enable simpler IT solutions that are easier and cheaper to deploy, integrate and upgrade, and to establish a standardized software model and market which benefits all parties (service providers, their suppliers and systems integrators).

Learn more about member collaboration

If you would like to learn more about the Open Digital Framework, or how to get involved in the TM Forum Collaboration Community, please contact George Glass.


the sustainable telco:
navigating the maze of scope 3 emissions

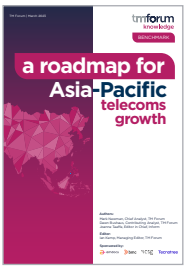

telco revenue growth:
time for operators to place new bets


Digital Transformation Tracker 7
cutting complexity
with automation and AI


making waves:
the future for Open RAN technology


partner ecosystems:
who does what and can CSPs compete for more?

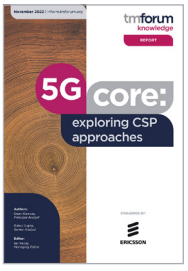

counter intelligence
using AI to improve customer experience


a roadmap for Asia-Pacific telecoms growth


establishing links:
platform models in the Open API economy


the sustainable telco:
engineering networks for net zero

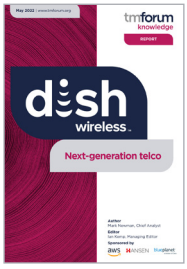

evolving
business support systems
for future services


5G core:
exploring CSP approaches

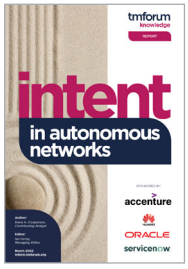

telco revenue growth:
taking it to the next level

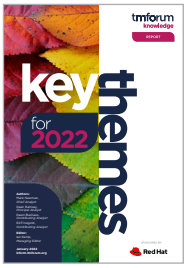

MEC: operators on the edge


Digital Transformation Tracker 6
The search for value intensifies

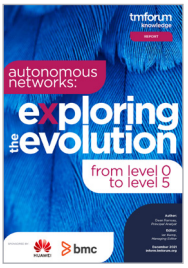

dish wireless.
Next-generation telco


telco to techco
capex and opex implications


intent in autonomous networks


key themes for 2022


improving customer experience in b2b markets


autonomous networks:
exploring the evolution
from level 0 to level 5


lessons learned
on the journey to cloud native


Cybersecurity strategies:
risk management moves firmly into the telco spotlight

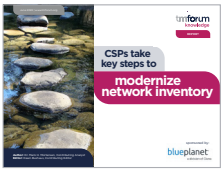

leveling up:
achieving Level 3 autonomous networks and beyond


wholesale changes:
rethinking support systems for new fiber operators


Reigniting telecoms growth
a Playbook for CEOs


how software-as-a-service is reshaping business support systems


transforming BSS:
racing to a flexible, customer-focused future

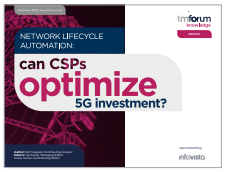

CSPs take key steps to
modernize network inventory


standout strategies:
how telcos are innovating in a crowded market


IoT at the edge:
taking opportunities beyond connectivity


mainframe modernization:
charting a course to cloud native


DIGITAL OPERATIONS MATURITY:
achieving business value
from transformation


NETWORK LIFECYCLE AUTOMATION:
can CSPs optimize 5G investment?


microservices:
paving CSPs' path to the cloud


autonomous networks:
from concept to reality


into the future:
can telcos monetize the metaverse?


DIGITAL ECOSYSTEMS
setting new standards for integration


how Liberty Latin America revamped to cloud BSS


best practice:
achieving level 3
autonomous operations


next generation
CEM now!

next generation
BSS
the Telefónica way

NaaS:
where it's going

reimagining operations & customer centricity

to connectivity and beyond!

process automation:
an evolutionary approach
to digital transformation

BSS strategies for
new lines of business

A balanced diet of
PUBLIC AND PRIVATE CLOUD

MOBILE PRIVATE NETWORKS:
Exploring the CSP opportunity

DevOps:
Where, why and how?

ORDER MANAGEMENT
FOR THE 5G ERA

AUTONOMOUS NETWORKS:
BUSINESS AND OPERATIONAL DRIVERS

tmforum knowledge

REPORT

meet the Research & Media team

Report Author:
Mark Newman
Chief Analyst
mnewman@tmforum.org

Practice Lead:
Dean Ramsay
dramsay@tmforum.org

Editor in Chief, Inform:
Joanne Taaffe
jtaaffe@tmforum.org

Head of Operations:
Ali Groves
agroves@tmforum.org

Sponsor Success Manager:
Maryssa Ramsey
mramsey@tmforum.org

Report Editor:
Ian Kemp
Managing Editor
ikemp@tmforum.org

Global Account Director:
Carine Vandevelde
cvandevelde@tmforum.org

Commercial Manager:
Tim Edwards
tedwards@tmforum.org

Digital Media Coordinator:
Maureen Adong
madong@tmforum.org

Marketing Manager:
Ritika Bhateja
rbhateja@tmforum.org

Report Design:
Intuitive Design UK Ltd
info@intuitive-design.co.uk

Published By:
European Office
TM Forum, Uncommon, 34-37 Liverpool Street,
London EC2M 7PP United Kingdom
Phone: +44 207 748 6615

US Office
181 New Road, Suite 304, Parsippany, NJ 07054 USA
Phone: +1 862-227-1648

ISBN: 978-1-955998-68-0

tmforum knowledge

REPORT

to learn more about TM Forum's work on autonomous networks and AI, please contact **Aaron Boasman-Patel**