ITUPublications

International Telecommunication Union
Standardization Sector

ITU-T Technical Report

(03/2025)

TR.GenAI-Telecom

Potential requirements and methodology for deploying and assessing generative AI models in telecom networks

Technical Report ITU-T TR.GenAI-Telecom

Potential requirements and methodology for deploying and assessing generative AI models in telecom networks

Summary

Generative artificial intelligence (GenAI) is a set of algorithms, capable of autonomously generating highly realistic content in various domains such as text, images, audio, or videos, and is an important solution to support accurate network modelling, network performance analysis, and automatic decision making for future networks including international mobile telecommunications (IMT)-2020 and beyond. Importantly, to realize the vision of telecom networks supported by artificial intelligence (AI), there is a clear need for a consolidated methodology for assessing and comparing GenAI models across a potentially large range of telecom use cases. This Technical Report studies the potential requirements of GenAI models that support telecom networks and evaluates their capabilities and performance in relevant use cases. This work lays the foundation for further studies conducted by international standards development organizations and industry alliances, paving the way for the integration of GenAI into telecom networks.

Keywords

AI, artificial intelligence, functional requirements, GenAI, generative artificial intelligence, performance requirements, telecom networks, telecom use cases.

Note

This is an informative ITU-T publication. Mandatory provisions, such as those found in ITU-T Recommendations, are outside the scope of this publication. This publication should only be referenced bibliographically in ITU-T Recommendations.

Table of Contents

Page

Technical Report ITU-T TR.GenAI-Telecom

Potential requirements and methodology for deploying and assessing generative AI models in telecom networks

1 Scope

This Technical Report studies potential requirements and methodology for deploying and assessing generative AI (GenAI) models in telecom networks.

The scope of this Technical Report includes:

- The study of the potential requirements on knowledge of telecom networks for GenAI;
- The overview of impactful use cases for GenAI in telecom networks, including potential requirements and implementation challenges, highlighting the expected benefits of GenAI for use cases, and the methodologies to assess and compare GenAI models with respect to the identified potential requirements;
- The study of potential functional requirements of telecom networks for supporting GenAI, the analysis of the potential risks of integrating GenAI to telecom operations, and the strategies to mitigate and assess the possible harmful impacts of GenAI.

2 References

This clause lists the ITU-T references cited in the Technical Report.

[ITU-T Y.3144]	Recommendation ITU-T Y.3144 (2024), *Future networks including IMT-2020 – Requirements and functional architecture of distributed core network.*
[ITU-T Y.3172]	Recommendation ITU-T Y.3172 (2019), *Architectural framework for machine learning in future networks including IMT-2020.*
[ITU-T Y.3187]	Recommendation ITU-T Y.3187 (2024), *Architectural framework for machine learning function orchestrator in future networks including IMT-2020.*
[ITU-T Y.3400]	Recommendation ITU-T Y.3400 (2023), *Coordination of networking and computing in IMT-2020 networks and beyond – Requirements.*
[ITU-T Y.3401]	Recommendation ITU-T Y.3401 (2024), *Coordination of networking and computing in IMT-2020 networks and beyond – Capability framework.*
[ITU-T CG-datasets]	Technical Report of ITU-T CG-datasets for AI/ML in networks (2024), *Datasets standardization approaches for datasets applicable for AI/ML in networks – First Edition,* <https://www.itu.int/md/T22-SG13-240715-TD-PLEN-0274/en>

3 Definitions

3.1 Terms defined elsewhere

This Technical Report uses the following terms defined elsewhere:

3.1.1 artificial intelligence (AI) [b-ETSI GR ENI 004]: Computerized system that uses cognition to understand information and solve problems.

NOTE 1 – [b-ISO/IEC 2382-28] defines AI as "an interdisciplinary field, usually regarded as a branch of computer science, dealing with models and systems for the performance of functions generally associated with human intelligence, such as reasoning and learning".

NOTE 2 – In computer science AI research is defined as the study of "intelligent agents": any device that perceives its environment and takes actions to achieve its goals.

NOTE 3 – This includes pattern recognition, the application of machine learning and related techniques.

NOTE 4 – Artificial-intelligence is the whole idea and concept of machines being able to carry out tasks in a way that mimics human intelligence and would be considered "smart".

3.1.2 generative artificial intelligence (GenAI) [b-ETSI GS ENI 030]: Type of artificial intelligence that can create new content (e.g., text, images or music) by learning the patterns and structures of existing data and then using those patterns to generate new data that is similar to the original data.

3.1.3 machine learning (ML) [ITU-T Y.3172]: Processes that enable computational systems to understand data and gain knowledge from it without necessarily being explicitly programmed.

3.2 Terms defined in this Technical Report

None.

4 Abbreviations and acronyms

This Technical Report uses the following abbreviations and acronyms:

AI	Artificial Intelligence
AN	Autonomous Networks
API	Application Programming Interface
CPU	Central Processing Unit
CSP	Communication Service Providers
CSV	Comma-Separated Values
GenAI	Generative Artificial Intelligence
GPT	Generative Pretrained Transformer
GPU	Graphics Processing Unit
IMT	International Mobile Telecommunications
IT	Information Technology
JSON	JavaScript Object Notation
LLM	Large Language Model
ML	Machine Learning
MoE	Mixture of Experts
RAG	Retrieval Augmented Generation
RAM	Random Access Memory
SQL	Structured Query Language
XML	Extensible Markup Language

5 Conventions

In this Technical Report, potential requirements which are derived from a given use case, are classified as follows:

The keywords "it is of critical value" indicate a possible requirement which would be necessary to be fulfilled (e.g., by an implementation) and enabled to provide the benefits of the use case.

The keywords "it is expected" indicate a possible requirement which would be important but not absolutely necessary to be fulfilled (e.g., by an implementation). Thus, this possible requirement would not need to be enabled to provide complete benefits of the use case.

The keywords "it is of added value" indicate a possible requirement which would be optional to be fulfilled (e.g., by an implementation), without implying any sense of importance regarding its fulfilment. Thus, this possible requirement would not need to be enabled to provide complete benefits of the use case.

6 Introduction

This clause presents the rationale for deploying and assessing generative artificial intelligence (GenAI) in telecom networks and provides an overview of GenAI models for telecom networks.

6.1 Rationale for deploying and assessing generative AI in telecom networks

In the ever-evolving telecommunications industry landscape, continuous innovation is imperative to meet the demands of a fully-connected and sustainable society. This innovation has paved the way for the birth of GenAI models with unprecedented abilities to acquire, process, and generate extensive volumes of comprehensive multimedia content. Among these remarkable strides, the generative pre-trained transformer series has emerged as a beacon, outshining its predecessors in both scale and capability.

As technology paves the way for ground-breaking advancements, GenAI models are capturing attention and interest across the telecom industry. A recent survey from a major consulting firm [b-Solon] highlighted that communication service providers (CSPs) are exploring GenAI for use cases that ameliorate customer services and network performance, e.g., from enhancing chatbot experience to improving network operations. Importantly, according to the survey, CSPs see GenAI as a distinct, significant incremental value on top of the "standard" AI/machine learning (ML), which is the focus of current activities in the standardization bodies.

The urgency for realizing the vision of telecom networks supported by GenAI is driving the attention of the information and communications technology industry worldwide. However, realizing this vision does not come without challenges and complexities. The primary concerns many stakeholders share are 1) the identification of use cases with the largest return on investment, 2) the capability of evaluating gains brought by GenAI with respect to standard expert-based or rule-based solutions, and 3) data security and governance.

To solve these challenges, standards organizations, industry alliances and open-source communities should join their efforts to develop a globally accepted methodology enabling the realization of the foreseen use cases. Some large language models (LLMs) use cases for telecom networks have been already introduced in the Technical Report of the ITU-T CG-datasets for AI/ML in networks "Datasets standardization approaches for datasets applicable for AI/ML in networks – First Edition" [ITU-T CG-datasets]; however, the potential requirements of these use cases are not investigated in the [ITU-T CG-datasets], which rather provides guidelines on datasets applicable for AI/ML in networks.

The selection of the most impactful use cases, from business and technology points of view, should be the results of a joint work from different stakeholders, within the telecom and AI ecosystems.

Importantly, to realize the vision of telecom networks supported by AI, there is a clear need for a consolidated methodology for assessing and comparing GenAI models across a potentially large range of telecom use cases.

Generally, the assessment of GenAI models should be based on the following three items:

- **Telecom knowledge evaluation**. Telecom knowledge evaluation refers to the definition of 1) the potential knowledge requirements from telecom GenAI models and 2) tests and related metrics to measure the telecom knowledge of the models. The expected knowledge for GenAI supporting network operations includes, but is not limited to, network architecture, protocols and functionalities. Clause 7 discusses the potential requirements of GenAI on knowledge of telecom networks and related integration methodology. Clause 10.2 presents how to assess telecom knowledge of GenAI models.
- **GenAI integration evaluation**. GenAI integration evaluation refers to the definition of 1) the potential risks of integrating GenAI models to telecom operations, and 2) strategies to mitigate and assess the harmful impacts of GenAI integration. The expected integration includes analysis of transparency, accountability, compliance, security and data privacy. Clause 11 discusses the potential risks of GenAI models in telecom networks and associated potential requirements.
- **Use case specific capability evaluation.** Use case specific capability evaluation refers to the definition of 1) the potential requirements in terms of capabilities (such as mathematical optimisation, software development, and text generation) from the perspective of GenAI models to support operations related to the telecom use cases, 2) the potential performance requirements (e.g., in terms of accuracy or energy consumption) to support GenAI operations related to the telecom use cases, and 3) a benchmarking methodology and related metrics to evaluate the capabilities of GenAI models. Clause 9 presents use cases of telecom operations where GenAI models are expected to provide a significant impact. Clause 12 presents potential requirements of GenAI for supporting specific telecom use cases.

The deployment of these items allows us to assess, compare, and improve GenAI models in telecom networks, which will accelerate the process of deploying GenAI models in telecom use cases.

6.2 Overview of generative AI models for telecom networks

AI relates to a computerized system that uses cognition to understand information and solve problems [b-ETSI GR ENI 004]. Such systems can solve simple problems such as automating tasks as well as very complex decision-making processes and have countless applications in various domains. The recent advances in AI have been transforming industries and are revolutionizing how people interact with technology and information.

A major disruptive force in this context is GenAI, a type of AI that can create new content (e.g., text, images or music) by learning the patterns and structures of existing data and then by using these patterns can generate new data that is similar to the original data [b-ETSI GS ENI 030]. GenAI usually leverages techniques like neural networks, to generate similar content that may seem human-generated material, sparking innovation and creativity. Within GenAI models, generative adversarial networks models [b-Goodfellow], in which two neural networks compete to improve their learned behaviours, and variational autoencoders models [b-Jain], which learn to compress data into a latent space and then reconstruct it, have both shown outstanding results in the fields of video and image generation.

Figure 1 describes the taxonomy proposed by [b-Gozalo-Brizuel], where nine categories of GenAI models are discussed: Text-to-image, Text-to-3D, Image-to-text, Text-to-video, Text-to-audio, Text-to-text, Text-to-code, Text-to-science, and other models, the last category including all the models that do not fit any of the previous categories, such as large action models [b-Masterman].

AI techniques are applied to countless fields. In telecommunications, they are used to analyse network data for traffic routing and failure predictions, enhancing network reliability, as well as enabling virtual assistants for customers such as real-time support or resolving queries.

GenAI presents unprecedented opportunities for innovation and optimization in the field. For instance, it can be used for generating real-life traffic for testing the quality or security of services. Moreover, LLMs can boost and make more efficient interactions with people, such as a more personalized interaction for customers or helping communication between people with different backgrounds and expertise.

NOTE – Despite its potential for transforming the telecom industry, GenAI is an incipient approach with unknown boundaries and costly requirements. Creating proper models usually requires massive amounts of data and computational resources, and the assessment of the uncharted social, economic, and technological impacts of their use is a major concern to the parties involved.

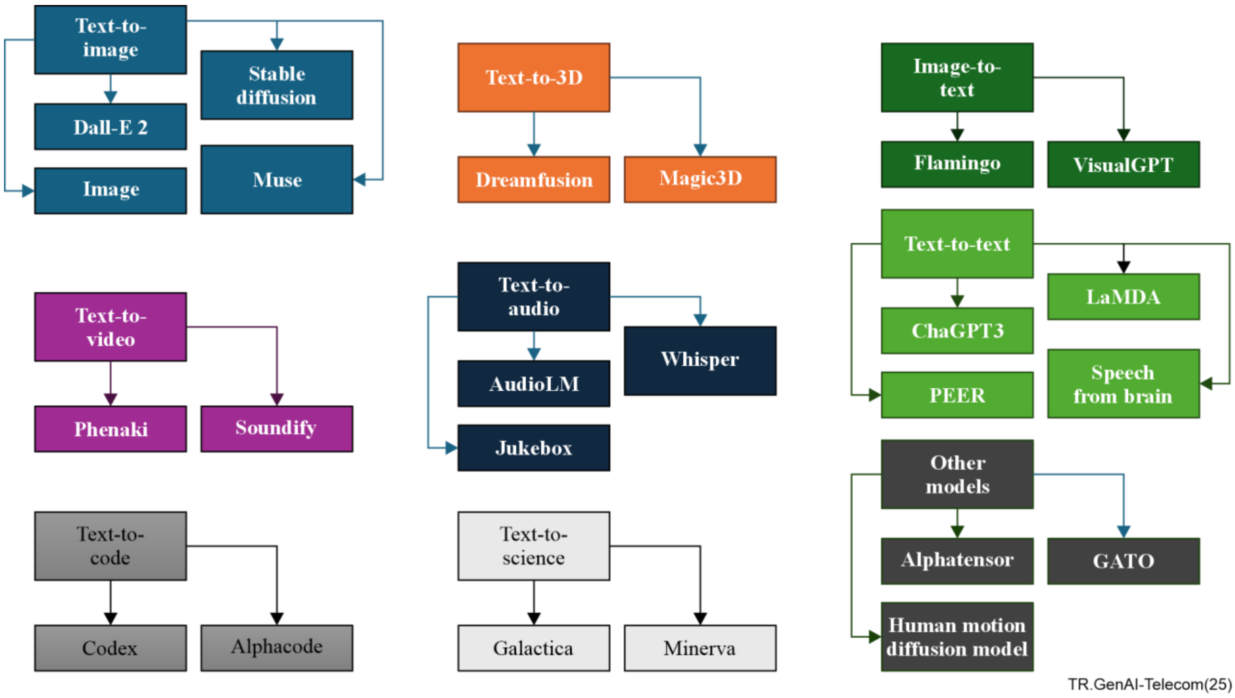

Figure 1 – A taxonomy of GenAI models proposed in [b-GC2023]

7 Potential requirements of generative AI on knowledge of telecom networks and related integration methodology

Telecom networks encompass various knowledge domains and GenAI is expected to properly understand and generate content aligned with real-world details and specific requirements of its applications [b-Zhou].

Although general-purpose GenAI models demonstrate considerable capabilities in various tasks, the diverse GenAI applications in telecom networks require domain-specific knowledge adaptation and use case specialized fine-tuning, to meet the expected performance requirements by improving accuracy, increasing user trust, and reducing hallucinations [b-Lewis]. An important design purpose of GenAI models is to serve as a foundation model for telecom networks. This requires a unified model trained for multiple different telecom tasks, like those described in clause 9. Therefore, adapting common telecom domain knowledge to general-purpose GenAI models is key to building generalized GenAI solutions for different telecom applications.

This Technical Report describes the potential requirements of GenAI models on knowledge of telecom networks in clause 7.1, and methods to integrate knowledge of telecom networks in GenAI models in clause 7.2.

7.1 Overview of potential requirements of Generative AI on knowledge of telecom networks

The different categories of telecom knowledge define the potential requirements of GenAI models on knowledge of telecom networks to support various telecom GenAI use cases:

- It is of critical value that GenAI models have knowledge of telecom network taxonomy and lexicon. For instance, LLMs for domain specific tasks provide better responses when trained on domain specific vocabulary [b-Soman]. The knowledge of domain specific terms facilitates clear alignment between the output of GenAI models and users' queries, as well as incorporates knowledge from the data.
- It is of critical value that GenAI models have knowledge of telecom standards: The telecom networks operate with protocols pre-defined in standards developing organizations, such as the International Organization for Standardization (ISO), Institute of Electrical and Electronics Engineers (IEEE), International Telecommunication Union Telecommunication Standardization Sector (ITU-T), European Telecommunications Standards Institute (ETSI) or the 3rd Generation Partnership Project (3GPP). The standard documents, including technical specifications and reports, specify the requirements and the capabilities of nowadays telecom systems, ranging from radio access networks, core networks, architecture and protocols, to service requirements. Incorporating telecom standard knowledge can support multiple GenAI models capabilities [b-Karapantelakis] [b-Bariah], such as: 1) Improve productivity for standard engineers on documentation; 2) Perform validation and testing of network products in compliance with standards; 3) Automate development of code t that implements standard features; 4) Analyse network performances in satisfying service requirements; 5) Simplify and streamline access to complex specifications, enhancing collaboration and understanding of industry standards [b-Lin, X].
- It is of critical value that GenAI models have knowledge of telecom research materials: the research papers, books and patents in the telecom domain provide a broader knowledge of the basic concepts, future evolution, and specialized solutions of telecom systems. For example, a research paper could contain mathematical modelling of different network scenarios, as well as several candidate algorithmic solutions to optimize the network towards a performance target. Such knowledge can assist GenAI models to perform as a network resource orchestrator [b-Maatouk], to analyse the operational status of a wireless system, to formulate the performance optimization goal into solvable problems, and to identify the appropriate tools to produce the related solutions. Moreover, the conceptual knowledge in the research materials can complement the technical specification language in standards, enabling GenAI models to produce more explainable responses to different use cases.
- It is of critical value that GenAI models have knowledge of telecom product implementations: the product specifications, code implementations, and application programming interface (API) documents can support GenAI in function development, performance testing and optimization for telecom products, prototypes, and simulators. This knowledge allows GenAI models to analyse the code in relation to standard specifications, and further realize the code implementing a feature or performing a system test as specified in human instructions [b-Nabeel]. Furthermore, the product knowledge allows GenAI models to assist network maintenance, e.g., finding the vulnerabilities of an implementation.
- It is of critical value that GenAI models have knowledge of telecom network operations: the operational logs, signaling messages, and network function configurations can enable GenAI models to support autonomous network configuration, troubleshooting and optimization. For example, as a network orchestrator, GenAI models can select the parameters with optimized settings to configure network functions [b-Bao]. GenAI models can analyse the system logs to determine optimization strategies, or generate network performance reports, as well as analyse the logs' semantic meaning for drafting trouble reports [b-Le]. Probe data for traffic

measurements enables traffic consumption to be accurately inferred, leading to better resource allocation [b-Zhang], and security focused applications can use API call sequence-based information for generating pseudo-benign malware to better train detection models [b-Peng] or, as in-phase and quadrature imbalance of transmitters, for generating fake signals [b-Roy].

- It is of critical value that GenAI models have knowledge of the telecom network environment: the network environment information, including user profiles, traffic patterns, network deployments and radio propagation scenarios, allows GenAI models to acquire a specific understanding of the telecom network. With this information, GenAI can predict the network state changes due to reconfiguration of network parameters, e.g., how the interference is affected by a transmit power adjustment. This allows the GenAI models to internally plan the optimal network configuration or command to achieve a performance target, before taking actions on the real network. Furthermore, the environment knowledge allows GenAI models to generate realistic synthetic radio environment or network traffic data, to assist testing of new network features [b-Anande].
- It is of critical value that GenAI models have knowledge of telecom customer support and experience: GenAI chatbots can assist customers with queries, troubleshooting, and providing personalized recommendations [b-Lin, X]. The information on the customers' experiences and journeys (i.e., brand interactions in pre-purchase, purchase and post-purchase stages) may be used not only to enhance GenAI models to provide more personalized services, but, also, to analyse how and where AI tools may be applied to improve in the journey's touchpoints [b-Moura]. GenAI models may also be used in churn predictions [b-Jain].
- It is of critical value that GenAI models incorporate risk-aware mechanisms to mitigate potential risks associated with knowledge of telecom networks, which poses challenges related to privacy, bias, transparency, compliance, and data sharing risks [b-Feuerriegel] [b-Solon]. In the network domain, data sharing presents a critical risk, and it is of critical value that data processing and storage for GenAI comply with regional and national regulations. To address these concerns, it is of critical value that GenAI models integrate mechanisms for risk assessment and mitigation, including real-time monitoring of model outputs, bias detection frameworks, secure data handling techniques, compliance verification modules, and strict data governance policies that regulate data localization and cross-border data transfers.

These requirements on telecom domain knowledge are complementary for GenAI models to perform different tasks. For example, when optimizing network configurations, GenAI should associate the parameters from the code defined in product implementations with their definitions in standards and should then utilize the knowledge of optimization methods described in the research literature to generate the best parameter settings. This requires GenAI models to be adapted to a mixture of the knowledge categories.

7.2 Methodologies to integrate the knowledge of telecom networks in generative AI models

This clause presents potential stages to build Telecom-specific GenAI models:

- Knowledge adaptation: the first step is to enhance the telecom domain knowledge on general purpose GenAI models. It can be achieved either through fine-tuning with a mixture of telecom and general domain data [b-Bariah], or through the retrieval augmented generation (RAG) from embedding database or knowledge graph of the telecom corpus [b-Ovadia].
- Instruction following: as a multi-task model, it is crucial to ensure GenAI models produce responses following the task specified in the instruction. This can be achieved through fine-

tuning with a mixture of multiple task relevant instruction data extracted from the telecom knowledge database [b-Ziegler].

- Policy alignment: to ensure GenAI models produce concise and accurate responses in low-latency and high-reliability telecom use cases, aligning GenAI output with task specific preferences is mandatory, which can be achieved through reinforcement learning [b-Sutton] on preference datasets.

NOTE – Preference datasets are used for reward modelling, where the downstream task is to fine-tune a baseline model in order to capture human preferences [b-PyTorch].

These stages not only enhance telecom domain knowledge of GenAI models, but also their capabilities in the telecom use cases. For instance, knowledge adaptation also reduces cost and latency inference from GenAI models.

8 Potential capabilities of generative AI agents and functional requirements of telecom networks for supporting generative AI agents integration

This clause discusses the potential capabilities of GenAI agents in telecom networks and the associated potential functional requirements of telecom networks for supporting GenAI agents' integration in telecom networks. These potential capabilities and requirements might extend some of those in [ITU-T Y.3144], with respect to AI/ML agents in a distributed core network, and some of those in [ITU-T Y.3400], with respect to coordination among resources of different types.

8.1 Potential capabilities of GenAI agents in telecom networks

In future telecom networks, AI agents [b-ETSI GR ENI 051] empowered by GenAI models (i.e., GenAI agents) can have the capabilities of taking human-like decision making processes, providing a path towards sophisticated and adaptive network protocols.

Also, GenAI agents can bring autonomy to the communication system, by having the capabilities of breaking down high-level business requirements into low-level actionable tasks and assigning them to different network elements for execution.

In addition, GenAI agents should be aware of network resource availability and capability of orchestrating network resources and controlling network functions in different domains of the telecom network through collaboration [ITU-T Y.3187] [ITU-T Y.3144]. This orchestration capability includes interfacing with network protocols in order to control e.g., telecom network and computing resources.

To support these awareness and orchestration capabilities in future networks empowered by GenAI agents, the following deployment architectures can be considered (see Figure 2):

- Hierarchical architecture: in this deployment architecture option, GenAI agents can operate in a hierarchical manner, with "leader" agents deployed in higher network layers and "worker" agents deployed in the underlaying data networks (e.g., access or core networks). The interactions between agents are hierarchical, where the leader agent, based on more capability and larger AI/ML models, plans and assigns tasks to the worker agents.
- Distributed architecture: in this deployment architecture option, the operations of the GenAI agents are not hierarchical. The interaction between these agents is self-organized and agents exchange past observations, actions, or strategies to collaboratively accomplish the tasks.

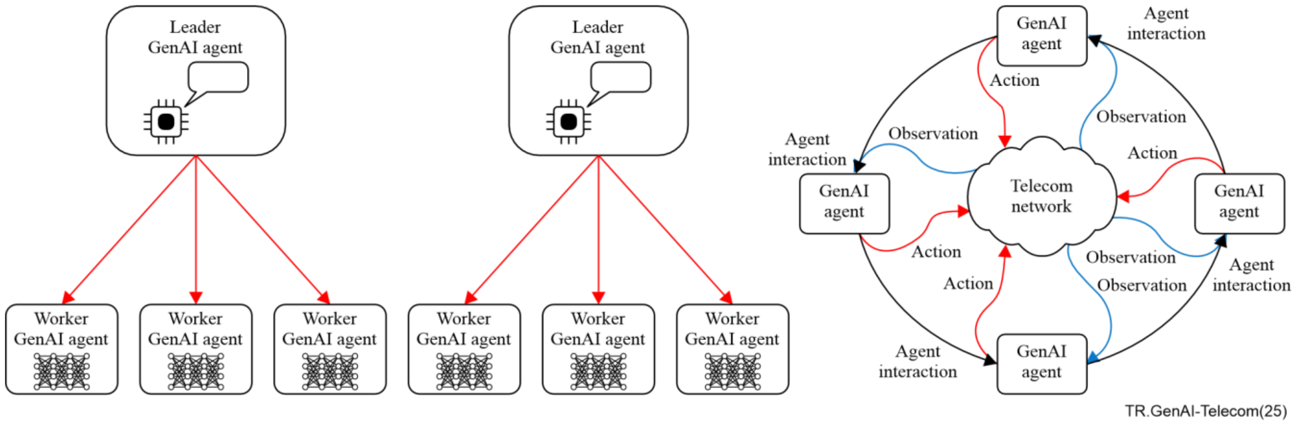

Figure 2 – Potential options of deployment architecture for GenAI agents, hierarchical architecture(left) versus distributed architecture (right)

NOTE 1 – Different types of AI-to-AI communication and their security risks and threats are discussed in [b-ETSI TR 104 031].

GenAI agents can also be leveraged to enhance network data quality, e.g., generating synthetic data, detecting and fixing inconsistencies in data. The data related to network operations or management can be collected from multiple sources, which may be heterogeneous or of difficult access. And the low number of connected devices, technical faults or network overload situations may lead to the scarcity of available data in the network. GenAI agents can generate synthetic sample data from available datasets, increasing data diversity and addressing data scarcity, especially in cases implying privacy and security concerns associated with data sharing, or high cost of data collection. In addition, GenAI agents can assist in identifying and fixing data inconsistency, providing clean and reliable datasets for model training, and improving model generalization.

To enable data augmentation capabilities empowered by GenAI agents in telecom networks, both deployment options of hierarchical architecture and distributed architecture can also be considered to support collaborative deployment and interaction between different GenAI agents in performing data augmentation tasks.

In addition to resource orchestration and control of network functions, a telecom network can use GenAI agents to bring autonomy to various network applications, such as autonomous vehicles and robots. The telecom network allows GenAI agents to exchange information in order to collaboratively perform various tasks, such as remote sensing, control, and planning.

NOTE 2 – GenAI capabilities can extend the existing framework for coordination of networking and computing [ITU-T Y.3401] by integrating further control capabilities. An example is a GenAI model embedded in an autonomous vehicle which is used to manage communication and computing resources together with the control of the vehicle driving system.

From the network application perspective, two types of GenAI deployment architecture can be considered too:

- Independent architecture: each network application is controlled by specific GenAI agents. In this case, for a network supporting autonomous vehicles, network-dedicated GenAI agents control network resources while application-dedicated GenAI agents control vehicles and traffic lights;
- Converged architecture: GenAI agents can jointly control network resources and applications. In this case, for a network supporting autonomous vehicles, GenAI agents deployed in autonomous vehicles can jointly control the vehicle and coordinate the communication with other vehicles, in order to collaboratively improve the traffic flow.

8.2 Potential functional requirements of telecom networks for GenAI agents integration

Based on the previously discussed GenAI capabilities and deployment architecture options, future telecom networks integrating GenAI agents should satisfy the following functional requirements:

- It is of critical value for telecom networks to expose to GenAI agents information on business intents;
- It is of critical value for telecom networks to expose to GenAI agents information on network operator's policies;
- It is of critical value for telecom networks to expose to GenAI agents information on network resource availabilities;
- It is of critical value for telecom networks to expose to GenAI agents information related to the applications supported by the telecom network;
- It is of critical value for telecom networks to expose information related to the performance of the GenAI agents running in the network;
- It is of critical value for telecom networks to support GenAI agents' capabilities for self-management and self-orchestration;
- It is of critical value for telecom networks to support GenAI agents' capabilities for controlling the applications supported by the telecom network;
- It is of critical value for telecom networks to support both distributed and hierarchical deployment architectures of GenAI agents;
- It is of critical value for telecom networks to enable information exchange between distinct GenAI agents to support collaborative orchestration of network resources and network functions;
- It is of critical value for telecom networks to enable GenAI agents to collaboratively control applications supported by the telecom network;
- It is of critical value for telecom networks to support the application, computing, and network resource coordination to enable their scheduling and management.

To support GenAI agents for data augmentation, future telecom networks should satisfy the following additional functional requirements:

- It is of critical value for telecom networks to support data collection for GenAI agents with different time granularities in terms of resources, networks and services;
- It is of critical value for telecom networks to support data processing for GenAI agents, including data processing to enhance data format, quality and security, e.g., by removing sensitive and private information;
- It is of critical value for telecom networks to support data storage for GenAI agents with the different architecture deployment options, enabling unified storage, access and management of telecom network related information;
- It is of critical value for telecom networks to enable the reuse of data exposed to GenAI agents, in order to avoid duplicate data collection and to minimize network overhead;
- It is of critical value for telecom networks to enable data lifecycle management of the entire data process for GenAI agents, including data collection, transmission, processing, storage and consumption.

NOTE – The potential requirements identified in this clause apply to both public networks and private networks supporting heterogeneous services and applications.

9 Use cases for generative AI in telecom networks

Although the extent of the impact of GenAI models in telecom networks in the following years is not measurable yet, several recently published studies have presented the expected and most relevant GenAI use cases. A recent survey, based on data from 104 senior-level respondents from 73 CSPs, has identified seven families of use cases which are either being explored already today, or have the short-to mid-term potential [b-Newman]: customer operations, sales and marketing, network, information technology (IT) and software engineering, product innovation, internal knowledge, and business operations. These family of use cases are illustrated in Table 1.

Table 1 – Families of GenAI use cases for CSPs [b-Newman]

Family	Use cases
Customer operations	Customer chatbot, call centre agent documentation and coaching, website assistance, predictive and personalized services
Sales and 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 and development	Evaluating new trends/developments, competitive analysis, supply chain analysis
Business operations	Contract, fraud management, partner management (e.g., roaming), human resources

More recently, an analysis has been published presenting how GenAI models could help telcos improve their revenues, based on a response from 130 telco operators in North America, Latin America, Europe, Africa, Asia, and the Middle East [b-McKinsey]. This document has identified five families of use cases: customer service, marketing and sales, network, IT and support functions (see Table 2, where the families are named "business domains" in line with the perspective described in clause 9.1).

NOTE – Overall, the families of use cases identified in the [b-Newman] and [b-McKinsey] surveys overlap, as the use cases on internal knowledge and business operations in [b-Newman] are related to those included in the support functions in [b-McKinsey]. It is interesting to notice that [b-Newman] also highlights the use cases on product innovation that stress the expected business value of GenAI models for telcos.

The following list provides more details related to the "network" family of use cases (see Table 1):

- Network planning: GenAI models can be utilized for network planning considering different objectives, e.g., coverage, capacity, bandwidth, etc. In addition, GenAI models can provide relevant planning tools to improve the efficiency of the planning process.
- Network deployment: GenAI models can be used for network deployment to reduce the launch time for new services or new features of the telecom networks, including the phases of configuration, testing and tuning.
- Network management and orchestration: GenAI models can be used to provide analysis and resolutions of network issues to assist management and orchestration of resources, functions and services, e.g., for coordination of networking and computing [ITU-T Y.3401], thus improving network reliability and stability.

9.1 Business value of the use cases

Table 2 describes the expected impact of GenAI models per business domain of use cases [b-McKinsey].

Table 2 – GenAI models impact on telcos by use cases [b-McKinsey]

Business domain	Share of total impact (%)	Share of surveyed business leaders focused on domain (%)	Example use cases
Customer service	35	85	Customer-facing chatbots, call-routing performance, agent copilots, bespoke invoice creation
Marketing and sales	35	45	Content generation, hyper-personalization, copilots for store personnel, customer sentiment analysis and synthesis
Network	15	62	Network inventory mapping, network optimization via customer sentiment analysis, self-healing via customer sentiment analysis on network problems
IT	10	55	Copilots for software development, synthetic data generation, code migration, IT support chatbots
Support functions	5	10	Procurement optimization, workplace productivity, internal knowledge management, content generation, HR Q&A

Table 2 highlights that more than 85 per cent of the executives' attribute to GenAI models more than 20 per cent of revenue or cost savings impact by business domain. Importantly, customer service together with marketing and sales, make up the largest share of the total expected impact in terms of business value. [b-McKinsey] reports that, for instance, in customer service, AI chatbots will improve customer support, which is anticipated to reduce its related costs by 15 to 20 per cent. Also, using GenAI models for summarizing voice and written client interactions is expected to reduce associated costs by up to 80 per cent.

In marketing and sales, CSPs can create GenAI models personalized messages and visual media to target individual customers. Using this tool, a European CSP reported a pilot project achieving more than a 10 per cent customer conversion rate. In the network family of use cases, GenAI models can improve network planning and management through e.g., the ability of analysing and structuring data about network components, including specifications and maintenance information from supplier contracts. In IT, software developers can complete coding tasks up to twice as fast with GenAI models [b-McKinsey Digital]. In support functions, GenAI models can reduce the costs associated with back-office operations and improve employee productivity: a European CSP anticipates that GenAI models will improve employee productivity by 30 per cent.

Similar business trends are also revealed in another recent report, based on interviews with more than 400 telecom professionals [b-NVIDIA]. According to it, among these respondents investing in AI, 57 per cent are using GenAI to improve customer service and support, 57 per cent to improve employee productivity, 48 per cent for network operations and management, 40 per cent for network planning and design, and 32 per cent for marketing content generation.

10 Methodology for assessment of generative AI models in telecom networks

This clause provides an overview of common methods to assess natural language processing and general (e.g., mathematics, history, law) knowledge of GenAI models. This clause also introduces a new approach to evaluate telecom knowledge of GenAI models.

10.1 Common methods for assessing generative AI models

Today there are three main ways to assess GenAI models: human evaluation, using a second model as a judge, and running a benchmark test using well established metrics [b-Zheng].

Using human labellers to judge the GenAI outputs is very time-consuming and costly. Also, this approach lacks flexibility as when the GenAI models or their tasks are updated, a new evaluation process is required.

Replacing the human judge with a GenAI model is promising as it reduces the cost and time constraints of human evaluation. However, the GenAI judge may not surpass human evaluation in accuracy and quality. A possible approach to test the GenAI judge is to create a small human evaluation dataset, which can test the accuracy of the GenAI judge.

As far as benchmark tests, Table 3 shows popular benchmark tests for assessing natural language processing and general (e.g., mathematics, history, law) knowledge of LLMs, i.e., GLUE (general language understanding evaluation) [b-Wang], SuperGLUE [b-Wang, A.], HellaSwag [b-Zellers], TruthfulQA [b-Lin], and MMLU (Massive multitask language understanding) [b-Hendrycks]. These standard benchmarks typically use well-known evaluation metrics, e.g., accuracy [b-Pedregosa] or F1 score [b-Pedregosa, F.] for classification tasks or question-answering tasks, and bilingual evaluation understudy (BLUE) [b-Papineni] or recall-oriented understudy for gisting evaluation (ROUGE) [b-Lin C-Y] for text generation tasks.

Table 3 – Popular benchmark tests for LLMs

Benchmark	Explanation	Metrics	Reference	URL
General language understanding evaluation (GLUE)	Standardized set of diverse natural language processing tasks	Correlation coefficients, accuracy, and F1 score	[b-Wang]	https://gluebenchmark.com/
SuperGLUE	More difficult language understanding tasks with respect to GLUE	Accuracy, exact match, and F1 score	[b-Wang, A.]	https://super.gluebenchmark.com/
HellaSwag	Benchmark for commonsense natural language inference	Accuracy	[b-Zellers]	https://github.com/rowanz/hellaswag
TruthfulQA	Benchmark made up of 817 questions designed to cause imitative falsehoods	Human evaluation	[b-Lin]	https://github.com/sylinrl/TruthfulQA
Measuring massive multitask language understanding (MMLU)	Multiple-choice questions related to 57 tasks including mathematics, history, computer science, and law	Accuracy	[b-Hendrycks]	https://github.com/hendrycks/test

10.2 Assessing telecom knowledge of generative AI models

TeleQnA is a benchmark dataset to assess the telecom knowledge of LLMs [b-Maatouk, A.]. This dataset is composed by 10 000 multiple-choice questions and answers related to different source materials, including resource publications, research overviews, standard specifications, standard overviews, and telecom lexicons. The fraction of the multiple-choice questions on the different topics is shown in Figure 3. TeleQnA uses two LLMs to generate and validate the questions and answers and integrates human-in-the-loop to verify the grammar of the questions and filter out duplicated and/or degenerated questions.

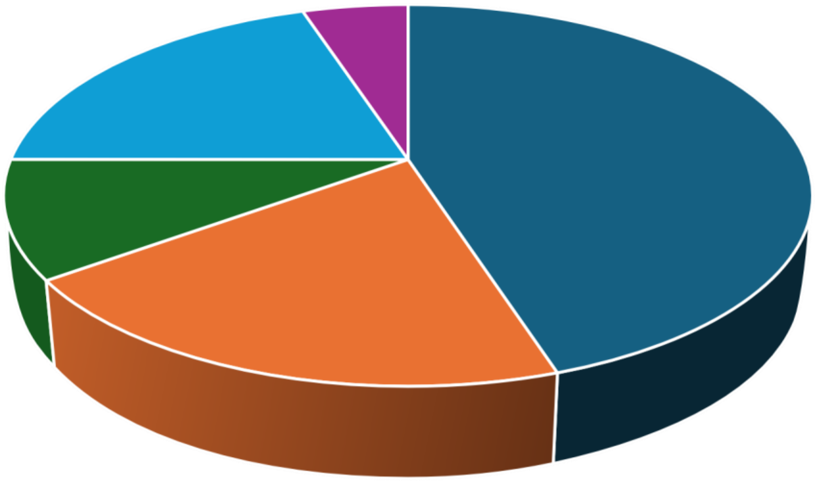

Figure 3 – Distribution of the TeleQnA dataset among the categories of the collected source materials [b-Maatouk, A.]

TeleQnA can be used to test the telecom knowledge of both general-purpose LLMs as well as specialized LLMs. Table 4 shows the accuracy results achieved by general-purpose LLMs on TeleQnA with respect to the categories of questions shown in Figure 3.

This analysis highlights that LLMs have comparable, and often better, capabilities to human experts in terms of telecom knowledge, underscoring the potential of LLMs within the telco domain. However, when looking at complex topics, such as questions related to standard specifications, LLMs performance is limited, which constraints their implementation to automate complex tasks in telecom networks. Therefore, to apply LLMs in telecom networks, a potential requirement is the integration of tools to specialize LLMs on telecom networks through e.g., RAG or fine-tuned models [b-Ovadia].

NOTE – These tools potentially introduce the need of knowledge databases, ML models, and new functionalities, able to process, store and retrieve relevant information to support the LLM operations.

Table 4 – Accuracy results of different LLMs and human experts on TeleQnA questions [b-Maatouk, A.]

Category of questions	Mistral 7b	Mixtral mixture of experts (MoE)	Generative pretrained transformer (GPT)-3.5	GPT-4	Humans
Lexicon (500)	56.8	83.8	82.2	86.8	80.33
Research overview (2000)	51.6	70.7	68.5	76.25	63.66
Research publications (4500)	49.27	70.2	70.42	77.62	68.33
Standards overview (1000)	46.2	66.73	64	74.4	61.66
Standards specifications (2000)	35.6	55.85	56.97	64.78	56.33

Table 4 – Accuracy results of different LLMs and human experts on TeleQnA questions [b-Maatouk, A.]

Category of questions	Mistral 7b	Mixtral mixture of experts (MoE)	Generative pretrained transformer (GPT)-3.5	GPT-4	Humans
Overall accuracy (10000)	47.07	67.74	67.29	74.91	64.86

11 Potential risks of generative AI models in telecom networks and associated potential requirements

GenAI models present significant challenges, such as ethical considerations, security vulnerabilities, and regulatory compliance [b-FOIS], [b-Feuerriegel]. Exploiting the GenAI benefits while avoiding the risks requires a comprehensive framework of technical standards, governance mechanisms, and risk assessment protocols [b-Solon] [b-European Union].

11.1 Performance risks and associated potential requirements

The most impressive GenAI models in use are built from learning on vast amounts of data. The current top performing models are trained with as much data as possible, usually terabytes of data (petabytes in the case of some LLMs). This not only impacts the necessary computing and storage infrastructure for creating and applying the models but also the performance and reliability of GenAI models which are directly affected by the quality and integrity of the datasets.

- It is of critical value to ensure the accuracy, completeness, and veracity of training data by assessing its quality and representativeness, pre-processing, and using cross validation.

GenAI models may be applied to network optimization, customer experience enhancement, service innovation, predictive maintenance, resource allocation optimization and more [b-Zhou].

- It is of critical value that the learning processes provide reliable GenAI models, behaving within expected boundaries as well as not generating incorrect results.
- It is expected to incorporate solutions to analyse the compatibility of GenAI models with current and legacy systems for interoperability and integration, as well as to continuously monitor GenAI for regulatory compliance, robustness, and reliability as well as data drifts.
- It is of critical value that user experience and usability are continuously assessed in the case of customer service.

11.2 Privacy risks and associated potential requirements

Training datasets usually have personal data information, an important issue for customers and regulators [b-Solon].

- It is of critical value that rules for obtaining, storing, and processing private data are well defined and in accordance with enforced regulations. This may be addressed by enforcing a proper governance framework and focusing on privacy by design approach, with regular compliance auditing.

11.3 Bias and discrimination risks and associated potential requirements

Synthetic data created by GenAI models may be used to provide artificial representations of real user data replicating statistical properties while safeguarding private data, as well as to augment existing datasets for training AI/ML models in order to improve their performance [b-Solon]. However, the generated data might replicate the biases present in the original datasets, reinforcing fairness and discrimination issues existing in GenAI models.

- It is of critical value that the data used for training is diverse and representative and that fairness-aware techniques are applied to mitigate biases.
- It is of critical value applying proper evaluation metrics to assess the quality, utility, and effectiveness of the synthetic data, and incorporating feedback mechanisms based on user feedback and domain knowledge.

11.4 Transparency and accountability risks and associated potential requirements

- It is of added value that telecommunication providers pursue research and development on the interpretability and explainability of GenAI models, looking into feature importance analysis and local/global explanation approaches [b-Feuerriegel]. Such efforts towards algorithmic transparency and accountability would positively impact customer trust and regulatory compliance.

11.5 Compliance risks and associated potential requirements

Incorporating GenAI models in telecommunications should observe regulatory frameworks on data protection and consumer rights since non-compliance may result in penalties, consumer mistrust and damage to CSPs [b-European Union]. Documentation and reporting along with continuous monitoring of the processes facilitate audits.

- It is of critical value that technical standards for evaluating the robustness, fairness, and transparency of the GenAI models are developed by international standards development organizations.

11.6 Security risks and associated potential requirements

Considering that the impacts of adopting GenAI models in telecommunications are not completely known yet, scenario analyses, vulnerability scans, and other assessments are important to handle risks and manage regulatory requirements [b-European Union], [b-Hacker].

- It is expected that functions and protocols are developed and integrated in the telecom network products for identifying, assessing, and mitigating the dangers associated with GenAI models, as well as for transparency and accountability.
- It is of added value to share knowledge and experiences among stakeholders in order to help managing risk scenarios.

As the importance of GenAI models in telecommunication systems increases, they also become a more and more critical point. The use of an additional layer of GenAI-powered software presents itself as another target for malicious actions aiming to disrupt the GenAI models and, thus, impact their performance [b-Campbell].

- It is expected to carefully consider security in data encryption, storage, and access in the network, and in network APIs, as well as enable awareness and continuous monitoring of the network status.
- It is expected to carefully design telecommunication systems avoiding being over reliant or completely dependent on GenAI-powered components.

11.7 Summary of potential risks of generative AI models in telecom networks and associated potential requirements

Table 5 summarizes the potential risks of GenAI models in telecom networks and associated potential requirements, which are detailed in the previous clauses.

Table 5 – Summary of potential risks of GenAI models in telecom networks and associated potential requirements

Type of potential risk	Potential risk	Potential requirements
Performance	• Performance of GenAI models depends on the quality and integrity of training datasets • The learning processes must provide reliable GenAI models, behaving within the expected results	• Training data should be accurate, complete and truthful • Continuously assess user experience, monitor GenAI models, and ensure compatibility with current and legacy systems
Privacy	Datasets may include personal data information	Rules for obtaining, storing and processing private data should follow enforced regulations
Bias and discrimination	Synthetic data used for training may be biased	• Integrate fairness-aware techniques in the data creation • Usage of assessment metrics • Incorporate feedback mechanisms
Transparency and accountability	Customer trust in GenAI models could be limited due to limited algorithmic transparency	Continue research and development on the interpretability and explainability of GenAI models
Compliance	Non-compliance on the regulatory framework may result in penalties, consumer mistrust and damage to the CSP	International standards development organizations should develop technical standards for evaluating the robustness, fairness, and transparency of the models
Security	Malicious actions aiming to disrupt the GenAI models may impact telecommunication systems Impacts of adopting GenAI models in telecommunications are not completely known yet	• Consider security in data encryption, storage, and access in the network and network APIs • Limit telecom network dependency on GenAI components • Realize scenario analyses, vulnerability scans • Develop functions and protocols for identifying, assessing, and mitigating the dangers associated with GenAI models

12 Potential requirements of GenAI specific for supporting telecom use cases

Telecom use cases have diverse and complex requirements spanning from energy consumption to knowledge of standard documents, which makes it challenging to identify the right GenAI model for each specific telecom use case. In this clause, the GenAI requirements specific for supporting the use cases for GenAI in telecom networks are introduced and separated into functional requirements and performance requirements.

12.1 Potential functional requirements

The following functional requirements apply for supporting the use cases for GenAI in telecom networks.

- It is of critical value that GenAI is capable of software development: GenAI should write, optimize and maintain software code, based on natural language descriptions, helping developers rapidly prototype or implement solutions. GenAI should also automate tasks such as debugging, refactoring, and providing suggestions for performance improvements. GenAI models should also be capable of generating code regular expressions, helping users create complex pattern-matching rules from natural language descriptions. Popular benchmarks to test GenAI capabilities in software development are HumanEval [b-Chen], massive benchmark for programming problems (MBPP) [b-Austin] and CodeXGLUE [b-Lu].

- It is of critical value that the output of the GenAI models complies with the expected format: GenAI models should generate content, whether code, text, or structured data, that adheres to specific formatting standards, conventions, or templates indicated by the user. Examples of relevant formats for telecom use cases are JavaScript object notation (JSON) and structured query language (SQL). A popular benchmark to test this capability is Spider [b-Yu].
- It is of critical value that GenAI supports diverse data structures as input: GenAI models should be able to process, interpret, and transform structured formats like JSON, extensible markup language (XML), comma-separated values (CSV), and database tables. By understanding the relationships within structured data, GenAI can perform tasks such as data extraction, transformation, validation, and even synthesis of structured data. A popular benchmark to test this capability is TableQAEval [b-Lei].
- It is of critical value that GenAI is capable of mathematical and logical reasoning: GenAI models should be able to realize mathematical and logical reasoning, allowing GenAI models to perform problem-solving, calculations, and algorithmic thinking. Popular benchmarks to test these capabilities are MATH [b-Hendrycks, D.] and FOLIO [b-Han].
- It is of critical value that GenAI supports tool calling: GenAI models should have tool invocation capabilities, enabling them to autonomously interact with various software tools, APIs and libraries based on natural language instructions. By recognizing the business intents, GenAI models can trigger specific actions – like fetching data, performing calculations, or automating tasks - through external tools without requiring direct manual input. A popular benchmark to test this capability is ToolBench [b-Qin].
- It is of critical value that GenAI supports capabilities for data augmentation.

12.2 Potential performance requirements

The following performance requirements apply for supporting the use cases for GenAI in telecom networks.

- It is of critical value that GenAI is sustainable: GenAI models demand substantial energy resources, especially during inference, which occurs continuously in telecom applications like customer service. Research indicates that inference consumes significantly more energy than training, as it runs in real-time and scales with user demand [b-Samsi].
- For CSPs aiming to optimize energy use, efficient benchmarking should consider the GenAI model architecture, deployment configuration, and optimizations such as batching and parallel processing [b-Hisaharo]. Recent studies reveal that hybrid graphics processing unit (GPU)-central processing unit (CPU) configurations can reduce energy use by up to 7.5%, underscoring the importance of adaptive infrastructure design [b-Wilkins].
- Although energy consumption is often secondary to performance in evaluations, there is a growing shift toward prioritizing energy efficiency to align with industry sustainability goals and reduce operational costs [b-Shi], as well as to match the societal requirements regarding carbon footprint. By implementing energy-aware strategies and GenAI infrastructure adaptations, CSPs can balance real-time performance with energy demands, ultimately supporting a more sustainable AI deployment strategy.

- It is of critical value that the latency of GenAI inference is constrained: the inference latency of the GenAI models should be limited to ensure effective service provisioning. In general, latency larger than 30s may lead to service disruptions.
- It is of critical value that the random-access memory (RAM) need of GenAI models is constrained: in scenarios where large document corpora are available for retrieval augmented generation (RAG), the volume of the embedded chunks can grow so large that it surpasses the RAM capacities of standard computing systems. This poses a significant challenge in terms of hardware requirements and processing efficiency.

Bibliography

[b-ETSI GR ENI 004] ETSI GR ENI 004 V3.1.1 (2023), *Experiential Networked Intelligence (ENI); Terminology*. <https://cdn.standards.iteh.ai/samples/64212/88fe9fbdb1a7400986e0f5606ef8c7cc/ETSI-GR-ENI-004-V3-1-1-2023-07-.pdf>

[b-ETSI GS ENI 030] ETSI GS ENI 030 V4.1.1 (2024), *Experiential Networked Intelligence (ENI); Transformer Architecture for Policy Translation*. <https://cdn.standards.iteh.ai/samples/63943/80071cdacb434d81bd5b289fe9392da3/ETSI-GS-ENI-030-V4-1-1-2024-03-.pdf>

[b-ETSI GR ENI 051] ETSI GR ENI 051 V4.1.1 (2025), *Experiential Networked Intelligence (ENI); Study on AI Agents based Next-generation Network Slicing*. <https://cdn.standards.iteh.ai/samples/72234/0a59409c208f4082a4d5166b7d777f32/ETSI-GR-ENI-051-V4-1-1-2025-02-.pdf>

[b-ETSI TR 104 031] ETSI TR 104 031 V1.1.1 (2024), *Securing Artificial Intelligence (SAI); Collaborative Artificial Intelligence*. <https://www.etsi.org/deliver/etsi_tr/104000_104099/104031/01.01.01_60/tr_104031v010101p.pdf>

[b-ISO/IEC 2382-28] ISO/IEC 2382-28:2015, *Information technology – Vocabulary*. <https://www.iso.org/standard/63598.html#:~:text=This%20standard%20of%20vocabulary%20taken,concepts%20relevant%20to%20this%20field.>

[b-Anande] Anande, T.J., Al-Saadi, S., and Leeson, M.S (2023), *Generative adversarial networks for network traffic feature generation*. <https://doi.org/10.1080/1206212X.2023.2191072>

[b-Austin] Austin, J., Odena, A., Nye, M., Bosma, M., Michalewski, H., Dohan, D., Jiang, E., Cai, C., Terry, M., Le, Q., and Sutton, C. (2021), *Program Synthesis with Large Language Models*, arXiv:2108.07732. <https://arxiv.org/abs/2108.07732>

[b-Bao] Bao, L., Liu, X., Wang, F., and Fang, B. (2019), *ACTGAN: Automatic Configuration Tuning for Software Systems with Generative Adversarial Networks*. <https://xinliu.engineering.ucdavis.edu/sites/g/files/dgvnsk9831/files/media/documents/ASE19Bao.pdf>

[b-Bariah] Bariah, L., Zou, H., Zhao, Q., Mouhouche, B., Bader, F., and Debbah, M. (2023), *Understanding Telecom Language Through Large Language Models*. <https://arxiv.org/abs/2306.07933>

[b-Chen] Chen, M., Tworek, J., Jun, H., Yuan, Q., Pinto, H.P.D.E, Kaplan, J., Edwards, H., Burda, Y., Joseph, N., Brockman, G., Ray, A., Puri, R., Krueger, G., Petrov, M., Khlaaf, H., Sastry, G., Mishkin, P., Chan, B., Gray, S., Ryder, N., Pavlov, M., Power, A., Kaiser, L., Bavarian, M., Winter, C., Tillet, P., Such, F.P., Cummings, D., Plappert, M., Chantzis, F., Barnes, E., Herbert-Voss, A., Guss, W.H., Nichol, A., Paino, A., Tezak, N., Tang, J., Babuschkin, I., Balaji, S., Jain, S., Saunders, W., Hesse, C., Carr, A.N., Leike, J., Achiam, J., Misra, V., Morikawa, E., Radford, A., Knight, M., Brundage, M., Murati, M., Mayer, K., Welinder, P., McGrew, B., Amodei, D., McCandlish, S., Sutskever, I., and Zaremba, W. (2021),

Evaluating Large Language Models Trained on Code, arXiv:2107.03374.
<https://arxiv.org/abs/2107.03374>

[b-Campbell] Campbell, M., and Jovanović, M. (2024), *Disinfecting AI: Mitigating Generative AI's Top Risks*.
<https://www.researchgate.net/publication/380295599_Disinfecting_AI_Mitigating_Generative_AIs_Top_Risks>

[b-European Union] European Union (2024), Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act).
<https://eur-lex.europa.eu/eli/reg/2024/1689/oj>

[b-FOIS] Federal Office for Information Security (2025), *Generative AI Models: Opportunities and Risks for Industry and Authorities*.
<https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/KI/Generative_AI_Models.html>

[b-Feuerriegel] Feuerriegel, S., Hartmann, J., Janiesch, C., and Zschech, P. (2023), *Generative AI*.
<http://dx.doi.org/10.2139/ssrn.4443189>

[b-Goodfellow] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio Y. (2020), *Generative adversarial networks*. Communications of the ACM, Volume 63, Issue 11, pages 139-144.
<https://doi.org/10.1145/3422622>

[b-Gozalo-Brizuel] Gozalo-Brizuela, R., and Garrido-Merchán, E.C. (2023), *ChatGPT is not all you need. A State of the Art Review of large Generative AI models*.
<https://arxiv.org/pdf/2301.04655>

[b-Hacker] Hacker, P., Engel, A., and Mauer, M. (2023), *Regulating ChatGPT and other Large Generative AI Models*.
<https://doi.org/10.1145/3593013.3594067>

[b-Han] Han, S., Schoelkopf, H., Zhao, Y., Qi, Z., Riddell M., Zhou, W., Coady, J., Peng, D., Qiao, Y., Benson, L., Sun, L., Wardle-Solano, A., Szabo, H., Zubova, E., Burtell, M., Fan, J., Liu, Y., Wong, B., Sailor, M., Ni, A., Nan, Li., Kasai, J., Yu, T., Zhang, R., Fabbri, A.R., Kryscinski, W., Yavuz, S., Liu, Y., Lin, X.V., Joty, S., Zhou, Y., Xiong, C., Ying, R., Cohan, A., and Radev, D. (2022), *FOLIO: Natural Language Reasoning with First-Order Logic*, arXiv:2209.00840.
<https://arxiv.org/abs/2209.00840>

[b-Hendrycks] Hendrycks, D., Burns, C., Basart, S., Zou, A., Mazeika, M., Song, D., Steinhardt, J. (2021), *Measuring massive multitask language understanding*.
<https://arxiv.org/abs/2009.03300>

[b-Hendrycks, D.] Hendrycks, D., Burns, C., Kadavath, S., Arora, A., Basart, S., Tang, E., Song, D., and Steinhardt, J. (2021), *Measuring Mathematical Problem Solving With the MATH Dataset*, arXiv: 2103.03874.
<https://arxiv.org/abs/2103.03874>

[b-Hisaharo] Hisaharo, S., Nishimura, Y., and Takahashi, A. (2024), *Optimizing LLM Inference Clusters for Enhanced Performance and Energy Efficiency*.
<https://www.techrxiv.org/users/812455/articles/1213926-optimizing-llm-inference-clusters-for-enhanced-performance-and-energy-efficiency>

[b-Jain] Jain, H., Khunteta, A., and Srivastava, S. (2022), *Telecom Churn Prediction Using CNN with Variational Autoencoder*. Smart Systems: Innovations in Computing. pp. 583 – 600.
<https://doi.org/10.1007/978-981-16-2877-1_55>

[b-Karapantelakis] Karapantelakis, A., Thakur, M., Nikou, A., Moradi, F., Orlog, C., Gaim, F., Holm, H., Nimara, D.D., and Huang, V. (2024), *Using Large Language Models to Understand Telecom Standards*.
<https://arxiv.org/abs/2404.02929>

[b-Le] Le, V-H., and Zhang, H. (2023), *Log Parsing with Prompt-based Few-shot Learning*.
<https://arxiv.org/abs/2302.07435>

[b-Lei] Lei, F., Luo, T., Yang, P., Liu, W., Liu, H., Lei, J., Huang, Y., Wei, Y., He, S., Zhao, J., and Liu, K. (2023), *TableQAKit: A Comprehensive and Practical Toolkit for Table-based Question Answering*.
<https://github.com/lfy79001/TableQAKit>

[b-Lewis] Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W-t, Rocktäschel, T., Riedel, S., Kiela, D. (2020), *Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks*.
<https://doi.org/10.5555/3495724.3496517>

[b-Lin] Lin. S, Hilton, J., and Evans, O. (2022), *TruthfulQA: Measuring How Models Mimic Human Falsehoods*, arXiv preprint arXiv:2109.07958.
<https://arxiv.org/abs/2109.07958#>

[b-Lin C-Y.] Lin, C-Y. (2004), *ROUGE: A Package for Automatic Evaluation of Summaries.*
<https://aclanthology.org/W04-1013.pdf>

[b-Lin, X] Lin, X., Kundu, L., Dick, C., Galdon, M.A.C, Vamaraju, J., Dutta, S., and Raman, V. (2024), *A Primer on Generative AI for Telecom: From Theory to Practice*.
<https://arxiv.org/abs/2408.09031>

[b-Lu] Lu, S., Guo, D., Ren, S., Huang, J., Svyatkovskiy, A., Blanco, A., Clement, C., Drain, D., Jiang, D., Tang, D., Li, G. (2021), *CodeXGLUE: A Machine Learning Benchmark Dataset for Code Understanding and Generation*, arXiv:2102.04664.
<https://arxiv.org/abs/2102.04664>

[b-Maatouk] Maatouk, A., Piovesan, N., Ayed, F., De Domenico, and A., Debbah, M. (2024), *Large Language Models for Telecom: Forthcoming Impact on the Industry*.
<https://arxiv.org/pdf/2308.06013>

[b-Maatouk, A.] Maatouk, A., Ayed, F., Piovesan, N., De Domenico, A., Debbah, M., Luo, Z-Q. (2023), *TeleQnA: A Benchmark Dataset to Assess Large Language Models Telecommunications Knowledge*. arXiv:2310.15051.
<https://arxiv.org/abs/2310.15051>

[b-Masterman] Masterman, T., Besen, S., Sawtell, M., Chao, A. (2024), *The Landscape of Emerging AI Agent Architectures for Reasoning, Planning, and Tool Calling: A Survey*, arXiv:2404.11584.
<https://arxiv.org/abs/2404.11584>

[b-McKinsey Digital] McKinsey Digital. (2023), *Unleashing developer productivity with generative AI.*
<https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/unleashing-developer-productivity-with-generative-ai>

[b-McKinsey] McKinsey & Company. (2024), *How generative AI could revitalize profitability for telcos.*
<https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/how-generative-ai-could-revitalize-profitability-for-telcos#/>

[b-Moura] Moura, S., Reis, J.L, and Rodrigues, L.S. (2021), *The Artificial Intelligence in the Personalisation of the Customer Journey – a literature review*.
<https://aisel.aisnet.org/capsi2021/28>

[b-NVIDIA] NVIDIA Survey Report (2024), *State of AI in Telecommunications*: 2024 Trends.
<https://resources.nvidia.com/en-us-ai-in-telco/state-of-ai-in-telco-2024-report?ncid=no-ncid>

[b-Nabeel] Nabeel, M., Nimara, D.D., and Zanouda, T. (2024), *Test Code Generation for Telecom Software Systems using Two-Stage Generative Model*.
<https://arxiv.org/abs/2404.09249>

[b-Newman] Newman, M. (2023), *Generative AI: Operators take their first steps*. TM Forum.
<https://inform.tmforum.org/research-and-analysis/reports/generative-ai-operators-take-their-first-steps>

[b-Ovadia] Ovadia, O., Brief, M., Mishaeli, M., and O. Elisha. (2024), *Fine-Tuning or Retrieval? Comparing Knowledge Injection in LLMs*.
<https://arxiv.org/abs/2312.05934>

[b-Papineni] Papineni, K., Roukos, S., Ward, T., and Zhu, W-J. (2002), *Bleu: a method for automatic evaluation of machine translation.*
<https://dl.acm.org/doi/10.3115/1073083.1073135>

[b-Pedregosa] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., and Thirion, B. (2011), *Scikit-learn: Machine learning in Python: Accuracy score.*
<https://scikit-learn.org/stable/modules/generated/sklearn.metrics.accuracy_score.html>

[b-Pedregosa, F.] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., and Thirion, B. (2011), *Scikit-learn: Machine learning in Python: f1_score*.
<https://scikit-learn.org/stable/modules/generated/sklearn.metrics.f1_score.html>

[b-Peng] Peng, X., Xian, H., Lu, Q., and Lu, X. (2021), *Semantics aware adversarial malware examples generation for black-box attacks*.
<https://doi.org/10.1016/j.asoc.2021.107506>

[b-PyTorch] PyTorch, *Preference Datasets*.
<https://docs.pytorch.org/torchtune/0.3/basics/preference_datasets.html>

[b-Qin] Qin, Y., Liang, S., Ye, Y., Zhu, K., Yan, L., Lu, Y., Lin, Y., Cong, X., Tang, X., Qian, B., Zhao, S., Hong, L., Tian, R., Xie, R., Zhou, J., Gerstein, M., Li, D., Liu, Z. and Sun, M. (2023), *ToolLLM: Facilitating Large Language Models to Master 16000+ Real-world APIs*, arXiv: 2307.16789.
<https://arxiv.org/abs/2307.16789>

[b-Roy] Roy, D., Mukherjee, T., Chatterjee, M., and Pasiliao, E. (2019), *Detection of Rogue RF Transmitters using Generative Adversarial Nets*.
<https://doi.org/10.1109/WCNC.2019.8885548>

[b-Samsi] Samsi, S., Zhao, D., McDonald, J., Li, B., Michaleas, A., Jones, M., Bergeron, W., Kepner, J., Tiwari, D., and Gadepally V. (2023), *From Words to Watts: Benchmarking the Energy Costs of Large Language Model Inference*.
<https://arxiv.org/pdf/2310.03003>

[b-Shi] Shi, J., Yang, Z., and Lo, D. (2024), *Efficient and Green Large Language Models for Software Engineering: Vision and the Road Ahead*.
<https://arxiv.org/pdf/2404.04566v1>

[b-Solon] Solon, A. (2023), *The opportunity in generative AI for Telecom*.
<https://pages.awscloud.com/GLOBAL-other-DL-generative-ai-for-telecom-whitepaper-2023-learn.html>

[b-Soman] Soman, S., and Ranjani, H., G. (2024), *Observations on LLMs for Telecom Domain: Capabilities and Limitations*.
<https://doi.org/10.1145/3639856.3639892>

[b-Sutton] Sutton, R.S., and Barto, A.G. (2018), *Reinforcement Learning, An Introduction, second edition*, The MIT Press.
<https://www.andrew.cmu.edu/course/10-703/textbook/BartoSutton.pdf>

[b-Wang, A.] Wang, A., Pruksachatkun, Y., Nangia, N., Singh, A., Michael, J., Hill, F., Levy, O., and Bowman, S.R. (2019), *SuperGLUE: A stickier benchmark for general-purpose language understanding systems*.
<https://arxiv.org/abs/1905.00537>

[b-Wang] Wang, A., Singh, A., Michael, J., Hill, F., Levy, O., Bowman, S.R. (2018), *GLUE: A multi-task benchmark and analysis platform for natural language understanding.* arXiv preprint arXiv:1804.07461.
<https://arxiv.org/abs/1804.07461>

[b-Wilkins] Wilkins, G., Keshav, S., and Mortier, R. (2024), *Hybrid Heterogeneous Clusters Can Lower the Energy Consumption of LLM Inference Workloads*.
<https://arxiv.org/abs/2407.00010>

[b-Yu] Yu, T., Zhang, R., Yang, K., Yasunaga, M., Wang, D., Li, Z., Ma, J., Li, I., Yao, Q., Roman, S., Zhang, Z., and Radev, D. (2018), *Spider: A Large-Scale Human-Labeled Dataset for Complex and Cross-Domain Semantic Parsing and Text-to-SQL Task.*
<https://arxiv.org/abs/1809.08887>

[b-Zellers] Zellers, R., Holtzman, A., Bisk, Y., Farhadi, A., Choi, Y. (2019), *HellaSwag: Can a machine really finish your sentence?* arXiv preprint arXiv:1905.07830.
<https://arxiv.org/abs/1905.07830>

[b-Zhang] Zhang, C., Ouyang, X., and Patras, P. (2017), *ZipNet-GAN: Inferring Fine-grained Mobile Traffic Patterns via a Generative Adversarial Neural Network*.
<https://doi.org/10.1145/3143361.3143393>

[b-Zheng] Zheng, L., Chiang, W-L., Sheng, Y., Zhuang, S., Wu, Z., Zhuang, Y., Lin, Z., Li, Z., Li, D., Xing, E.P., Zhang, H., Gonzalez, J.E., and Stoica, I. (2023), *Judging LLM-as-a-Judge with MT-Bench and Chatbot Arena*. arXiv:2306.05685.
<https://arxiv.org/abs/2306.05685>

[b-Zhou] Zhou, H., Chengming, Hu., Yuan, Y., Cui, Y., Jin, Y., Chen, C., Wu, H., Yuan, D., Jiang, L., Wu, D., Liu, X., Zhang, C., Wang, X., and Liu., J. (2024), *Large Language Model (LLM) for Telecommunications: A Comprehensive Survey on Principles, Key Techniques, and Opportunities*.
<https://arxiv.org/abs/2405.10825>

[b-Ziegler] Ziegler, D.M., Stiennon, N., Wu, J., Brown, T.B., Radford, A., Amodei, D., Christiano, P., and Irving, G. (2020), *Fine-Tuning Language Models from Human Preferences*.
<https://arxiv.org/abs/1909.08593>