Large Language Models for Telecom: The Next Big Thing?

Lina Bariah $^1$, Qiyang Zhao $^2$, Hang Zou $^2$, Yu Tian $^2$, Faouzi Bader $^2$, and Merouane Debbah $^2$

$^1$Technology Innovation Institute
$^2$Affiliation not available

October 31, 2023

Abstract

The evolution of generative artificial intelligence (GenAI) constitutes a turning point in reshaping the future of technology in different aspects. Wireless networks in particular, with the blooming of self-evolving networks, represent a rich field for exploiting GenAI and reaping several benefits that can fundamentally change the way how wireless networks are designed and operated nowadays. To be specific, large language models (LLMs), a subfield of GenAI, are envisioned to open up a new era of autonomous wireless networks, in which a multimodal large model trained over various Telecom data, can be fine-tuned to perform several downstream tasks, eliminating the need for dedicated AI models for each task and paving the way for the realization of artificial general intelligence (AGI)-empowered wireless networks. In this article, we aim to unfold the opportunities that can be reaped from integrating LLMs into the Telecom domain. In particular, we aim to put a forward-looking vision on a new realm of possibilities and applications of LLMs in future wireless networks, defining directions for designing, training, testing, and deploying Telecom LLMs, and reveal insights on the associated theoretical and practical challenges.
Large Language Models for Telecom: The Next Big Thing?

Lina Bariah, Qiyang Zhao, Hang Zou, Yu Tian, Faouzi Bader, and Merouane Debbah

Abstract—The evolution of generative artificial intelligence (GenAI) constitutes a turning point in reshaping the future of technology in different aspects. Wireless networks in particular, with the blooming of self-evolving networks, represent a rich field for exploiting GenAI and reaping several benefits that can fundamentally change the way how wireless networks are designed and operated nowadays. To be specific, large language models (LLMs), a subfield of GenAI, are envisioned to open up a new era of autonomous wireless networks, in which a multimodal large model trained over various Telecom data, can be fine-tuned to perform several downstream tasks, eliminating the need for dedicated AI models for each task and paving the way for the realization of artificial general intelligence (AGI)-empowered wireless networks. In this article, we aim to unfold the opportunities that can be reaped from integrating LLMs into the Telecom domain. In particular, we aim to put a forward-looking vision on a new realm of possibilities and applications of LLMs in future wireless networks, defining directions for designing, training, testing, and deploying Telecom LLMs, and reveal insights on the associated theoretical and practical challenges.

I. INTRODUCTION

While recent wireless networks have witnessed several revolutions in terms of technological trends, it is apparent that future wireless generations are converging towards solidifying the principle of self-built, self-evolving networks, particularly with the maturing of the self-organizing networks (SON) paradigm and the remarkable advancements in artificial intelligence (AI) technologies [1]. The core principle of SON is based on the concept of enabling wireless networks to have the capability to adjust, reconfigure, and optimize their functions and parameters according to particular network conditions and user demands. However, despite the fact that SON is aimed for realizing automation in wireless networks, its performance is dependent on predefined network conditions and their corresponding configurations. The ultimate vision of self-driven networks is to realize a retained network performance, maintain sustainable resilience to network variations, and design versatile networks that are capable of handling new network conditions and scenarios.

As a cornerstone, generative AI (GenAI) can be a key player in realizing the vision of self-evolving networks. Within this context, Large Language Models (LLMs), a subfield of natural language processing and GenAI, have recently attracted a considerable attention from the research community as a revolutionary technology in the field of AI [2]. Among several LLMs, Falcon LLM [3], generative pre-trained transformer (GPT)-2/3/4, Bidirectional Encoder Representation from Transformer (BERT), large language model Meta AI (LLaMA), as well as visual generative models, e.g., DALL-E and Contrastive Language-Image Pre-Training (CLIP), have strongly impacted how AI is employed for inference and decision-making purposes, and laid down a new base for novel applications that can exploit the potential of GenAI models. This is rooted to the generative and predictability capabilities of these LLMs, in which large models (mainly based on the transformer architecture) are trained over a vast amount of unlabeled multimodal data (primarily textual and/or visual data), and therefore, are enabled to understand and generate human-like languages. Through the self-attention mechanism of transformers and the large amount of training data, the developed large models are able to capture the statistical patterns and relationships in the provided data, and hence, to predict and generate the required data. Similar approach applies to visual models, where variational autoencoders (VAEs) and generative adversarial networks (GANs) can be leveraged to map contextual data with images and vice versa.

Being a promising candidate to revolutionize several technologies in various fields, we believe that GenAI models can introduce a radical change in wireless network design and operation. In particular, we anticipate that LLMs will introduce a tangible enhancement in the performance of different schemes within the Telecom domain. This can be achieved through exploiting the generative capabilities of LLMs in addition to the multimodality nature of the data acquired in wireless networks, including radio frequency (RF) signals, and 2D and 3D visual representations of wireless environments, to attain improved contextual, situational, and temporal awareness, and therefore, enhanced wireless communication. Furthermore, it is noted that LLMs can enable wireless networks to enjoy predictability features, and hence, realize improved and proactive localization, beamforming, power allocation, handover, as well as, spectrum management, even for unseen network scenarios. On the other hand, wireless networks can contribute to the development of improved GenAI models, by allowing multiple machines to communicate reliably through the utilization of LLMs.

Over the last couple of decades, AI has been playing a key role in a wide range of applications in Telecom networks. However, most AI solutions in wireless networks are developed for solving dedicated problems, which is inefficient for general use in wireless network with emerging applications. LLMs on the other hand, represent a promising paradigm to a native AI network, which aims to solve many downstream and upstream tasks from a pre-trained model on network data, such as radio signals, network traffic, and system specification. We foresee that LLM-like solutions could generalize over unseen data, tasks, and scenarios, and reduce the complexity, cost, and power in network design, deployment and operation. To
achieve native AI networks, LLM should play a position in the following two aspects:

- **LLM for Wireless:** LLM is expected to enable multiple tasks in wireless communications and sensing (Fig. 1). For example, a pre-trained LLM on radio signal can potentially perform channel estimation, beamforming, power allocation, modulation and coding, and so on. Furthermore, multi-modal LLM could associate radio signal with image, point-cloud, or sound, which can enhance many applications such as environment reconstruction, positioning, and human pose estimation.

- **Wireless for LLM:** As future network devices equipped with LLMs, 6G networks should empower collective intelligence with new communication paradigms. Wireless networks will be transferred from data and model based to knowledge and reasoning based. In doing so, multiple on-device LLMs can interact with each other to accomplish complex tasks with limited resource and energy.

A. Contribution

While there are several articles that discussed foundation models in general, and LLMs in particular, e.g., [4]–[7] - and the references therein, all of these articles have focused on the application of LLMs in the general domain, with specific interest in textual data with minor discussions on visual data. In this article, *for the first time in the literature*, we approach LLMs from a Telecom perspective, where we explore the integral role that LLMs will play in future wireless networks. In particular, we aim to unveil how LLMs can be used in the Telecom industry, and we outline the opportunities that can be offered by LLMs for improved wireless sensing and transmission. Alternatively, we draw a roadmap to how wireless networks can potentially contribute to developing efficient large models, that are capable of accommodating the needs of future 6G networks. We finally reveal how the two paradigms (LLM for wireless and wireless for LLM) pave the way to the conceptualization of the artificial general intelligence (AGI)-empowered wireless networks, the seed to fully self-evolving networks.

II. LLM FOR WIRELESS

A. Large Language Models for Sensing

1) **3D Wireless Imaging:** Deep Learning (DL) models have fundamentally contributed to the development of improved wireless sensing schemes, in which RF data can be acquired and mapped to 2D images for sensing applications, including localization, remote sensing, and resource allocation. While DL approaches have demonstrated an acceptable performance in several scenarios, they lack the generalizability to new network requirements and scenarios, as well as the complexity arises from the reliance on supervised training data. For the latter, to achieve high-resolution sensing for mission-critical applications, large labeled data-sets are required, rendering the labeling process a huge burden. Within the same concept, neural radio-frequency radiance fields (NeRF) framework [8] was proposed to reconstruct a 3D images from the acquired RF data, relying on the DL architecture. NeRF demonstrated a promising performance in localization applications, however, it suffers from high computational complexity and lacks the ability to scale. The recent evolution in visual GenAI models has opened the horizon to a new era of sensing capabilities, in which machines now have the potential to generate high-quality images and understand visual content, i.e., generating 2D and 3D images from textual description or mapping images with corresponding text.

Among others, DALL-E (capable of generating high-quality images from textual descriptions), CLIP (capable of generating textual descriptions of images), have been identified as core models for 2D and 3D image generation from text, or vice versa. These models rely on the several advanced techniques including GANs and VAEs to capture the visual patterns and their contextual relationship with textual content.

The promising potentials of visual GenAI models introduce a plethora of benefits that can be exploited for wireless communications, for enhanced system design and optimization. This vision is built upon the principle of enabling GenAI models to understanding the cross features between different data modalities, including images of wireless propagation environments and RF signals. Specifically, these visual models will be developed to generate super-resolution 3D images of the surrounding environment from measured wireless data. The constructed 3D image will be then leveraged to enable improved communication schemes, in which enhanced contextual and situational awareness can be realized, and therefore, achieving improved beamforming, handover, resource allocation, etc. This is due to the fact that Telecom visual GenAI models are anticipated to capture the relationships between RF data and different information that can be extracted from 3D images (static system topology, non-moving objects, dynamic objects information), and the impact of the latter on the former, enabling accurate multi-dimensional reconstruction of the wireless environment from the acquired RF signals. Unlike conventional visual GenAI models that map images with texts using text/image encoders/decoders, the architecture of Telecom visual GenAI models are aimed to comprise encoders that are capable of extracting features and patterns from wireless signals and map these features into 3D images. Telecom-oriented LLM framework can be developed to extract and concatenate images and signals embeddings, and allow the machine to understand the relationship between the two data modality through self-attention mechanism.

2) **Super-Resolution Localization:** Super-resolution localization is an essential element in current wireless networks, since determining the accurate locations of different wireless devices or nodes in the network readily facilitate the design and optimization of different network functionalities, in order to accommodate the needs of different users according to their current status and positions. Current localization techniques follow mainly two directions, RF-based localization, which rely on RF signals from anchor nodes for position detection, or vision-aided localization, which exploits visual data and image processing for feature extraction and object detection.
While the latter can provide more contextual information that render higher localization accuracy, the limited field-of-view and calibration and precise alignment of cameras or sensors constitute limiting factors in such approaches. These challenges are further pronounced in multimodal localization, in which the fusion and alignment of different data modalities results represent a bottleneck in exploiting the full potential of multimodal localization. However, it is strongly believed that multimodality is essential for capturing contextual information, as well as users’ behaviours and activities, and hence, realizing super-resolution localization accuracy.

Within this context, LLM can be a game changer in enabling efficient multimodal localization schemes, in which the general and self-attention nature of these large models can be the key to detect the contextual and situational information of network users and nodes, capture the mutual displacement between multiple images, and cross-correlate these images and their variations with the corresponding electromagnetic behaviour of wireless signals. The benefits of utilizing LLMs for high-precision localization are two-fold. First, LLMs can introduce improved perception skills of the surrounding environment through integrating multimodal data that allows the understanding of the environmental, temporal, and situational aspects of the network events and behaviours, and their impact on the network performance. With the generative capabilities of LLMs, it is envisioned that a pretrained large model will be capable of accurately locating multiple users within the network, when the model receives 3D environmental information with measure RF data. Second, the exploitation of LLMs in localization applications is stemmed from the fact that pretrained large models allow wireless networks to enjoy improved predictability skills. In more details, the integration of 3D images with RF data enables the large models to identify both idle and active users, as well as predict their future activities through their generative and prediction capabilities. This offers a comprehensive understanding of the wireless environment and its variations, and accordingly, enable enhanced network configuration and optimization. This yield more proactive networks, paving the way to self-built and self-healing networks.

B. Large Language Models for Transmission

1) Multi-modal beamforming: The evolution of high-frequency communications, e.g., mmWave and THz communication, has given a rise to several issues pertinent to the transceiver design, as well as the reliability of signals transmitted over high frequencies. This is due to the lossy and highly directional nature of high-frequency communication, which yields severe signals degradation even with minor blockage or beam misalignment. Accordingly, efficient beamforming and beam alignment mechanisms are essential to be developed to achieve reliable high-frequency communication. Typically, beam selection is performed through the reliance on a predefined beam codebook, in which a wide-range of beam sweeping is performed between the transmitter and the receiver in order to select the optimum beam to establish a reliable link between the two nodes. Such a procedure is usually associated with high beam training overhead, introducing a performance limiting factor in high-mobility and latency-sensitive applications.

LLMs, pretrained over a large data-set of beamforming scenarios, have the capability to predict the optimum beam that maximizes the signal strength and minimizes interference. This can be achieved through exploiting multimodality for providing additional information in regard to blockage probability and users status and activities, and hence, an LLM can be used to capture the various features and patterns within the network dynamicity, and predict the optimum beam according to current and future network scenarios.

2) Frequency Division Duplexing (FDD) Transmission: Within the context of FDD, LLMs can be utilized for channel state information (CSI) estimation purposes between the uplink and downlink transmissions. In conventional FDD systems, where separate frequency bands are allocated for uplink and downlink transmissions, CSI acquisition is typically performed separately for the two directions, consuming the network resources and introducing high latency. This issue is more pronounced in massive multiple-input multiple-output (MIMO) scenarios, where it is very challenging to acquire all CSI over the uplink [9]. Alternatively, partial uplink CSI knowledge can be exploited to extrapolate the full downlink CSI. Through the self-attention mechanism and the generative capabilities of LLMs, we envisage that LLMs will be able to capture the inherent relationship between the uplink and down-
link transmission and exploit 3D multimodal environment data (including camera, radar, LiDAR, and GPS) in order to select the optimum uplink and downlink beam pair that yields perfect alignment between the angle-of-arrival and angle-of-departure, at a particular user position (Fig. 2). It should be emphasized that super-resolution localization achieved by LLMs (as discussed in Sec. II-A) can directly impact the beamforming performance in FDD systems, in which accurate localization allow LLM-enabled beamforming schemes to predict future user position, and therefore, enable improved planning for the spectrum resource over the uplink and downlink transmissions.

3) Joint Source-Channel Coding (JSCC): Conventional wireless networks are designed in a block-based architecture, where each process has a dedicated block, including channel estimation, coding, decoding, equalization, to name a few. With the identified requirements of future 6G networks, such architecture cannot achieve the envisioned key performance indicators of latency, reliability, spectral and energy efficiency, and connectivity. It is further emphasized that block-based architectures are difficult to scale to large, more generalized scenarios [10]. In this regards, current research initiatives have been conducted to explore the appealing benefits of integrating channel and source coding into semantic-aware JSCC, in order to serve use-cases with extreme latency and bandwidth requirements, that are computationally demanding in long block-length techniques. When incorporating different data modality, e.g., images and videos, to be transmitted, DL approaches have demonstrated outstanding performance in terms of latency and quality-of-compression. However, DL models rely heavily on a huge amount of labeled training data for the joint optimization of source and channel coding, which, in most scenarios, is not available. This readily impact the generalizability of the trained models to unseen or varying (channel conditions, interference, noise, etc.) network scenarios.

In this regard, LLMs, can facilitate the realization of efficient JSCC schemes for improved wireless communication. In addition to the labeled data independency, LLMs can be exploited to understand the statistical behavior of the source data and to allow the reliable extraction of the needed information, and hence, realize efficient data compression. Furthermore, through the self-attention mechanism, LLM introduce improved robustness to channel errors, through enabling efficient error correction mechanisms. This can be achieved by using LLMs to capture the inherent features of wireless channels behavior, error statistics, and source data, and then learn the relation between the source data and the corresponding channel coding requirements, and therefore, realize improved prediction for errors and promotes the development of robust error correction codes. Moreover, by understanding the long interdependencies between the source data and channel coding, adaptive mechanisms can be designed with reliance on LLMs, where real-time adjustment of the code rate, modulation, and code selection can be performed according to the current channel condition, striking a balance between source coding and channel coding performance.

III. WIRELESS FOR LLM

A. 6G with Collective Intelligence

An evolution of GenAI is to empower a massive number of wireless devices to deliver collective intelligence. To achieve this, communication systems should be transferred from data and model-based to knowledge and reasoning-based paradigms. Conventional communication networks aim to transfer data from one network node to another under targeted key performance indicators (KPIs). This is inefficient for a network connecting large-scale devices powered by LLMs. The goal of 6G is to develop a computing fabric that moves knowledge within the network. This can push nowadays cloud LLM trained on world knowledge towards distributed collective intelligence. To achieve this, LLMs should be grounded to the real-world context, communicate with knowledge, to perform multi-agent planning, decision making, and reasoning.

System 2 Machine Learning (ML) includes the capability of grounding, planning, criticising, and reasoning. This becomes more important for on-device LLM where the cost of inference and fine-tuning is much higher than in the cloud. For an effective implementation, the LLM knowledge should be grounded to the real-world context, communicate with knowledge, to perform multi-agent planning, decision making, and reasoning.

8. System 2 Machine Learning (ML) includes the capability of grounding, planning, criticising, and reasoning. This becomes more important for on-device LLM where the cost of inference and fine-tuning is much higher than in the cloud. For an effective implementation, the LLM knowledge should be grounded to the real-world context, communicate with knowledge, to perform multi-agent planning, decision making, and reasoning.

Planning involves the capability of an LLM to create sequences of actions to accomplish a task. This can help wireless agents to distribute sub-tasks to different devices so as to improve the execution efficiency. As a generative model, an LLM can make wrong decisions and answers, where the criticism between multiple LLM instances is proved to improve the performance [12]. On the other hand, reasoning includes solving complex problems from priors and beliefs, such as chain or tree of thoughts [13]. By breaking down a large problem to small ones, LLMs can provide more effective inference, thus reducing the energy cost.

Semantic information and communication is a potential framework for knowledge and reasoning driven networks. LLMs can learn the semantic abstraction of data and represent it as knowledge. In this context, information should be characterized with a minimal structure, which is robust against changes in distribution, domain and context [14]. This does not only reduce the size of data and model in the device memory, but also can represent different data modalities with a

![Fig. 2: LLM for Beamforming in FDD Systems](image-url)
common latent (concept) space. Knowledge can be potentially modeled on a topological space, where LLMs can leverage lower-order information for short-term action, and higher-order abstractions for long-term plan. In this context, LLMs have the flexibility to encode the domain, task, or action specific knowledge, such that the communication between LLMs can be resource and energy efficient.

With the capability of planning and reasoning on semantic knowledge, multiple on-device LLMs should collaborate to complete the intended goal. Multi-agent reinforcement learning (MARL) is an approach to maximize a collaborative long-term goal (reward) from multiple agents, by learning a value function conditioned on the environment states, observed by the agents. MARL can be integrated with LLMs by rewarding its generated actions, to ground LLMs in a multi-agent scenario. Furthermore, game theory can be a potential technique for enabling the interactions between multiple LLMs in a competition scenario [15]. For example, the agents could initiate tasks with different goals, and perform actions under shared resources. Multiple LLMs could play a non-cooperative game to complete a common or several different tasks in a competition scenario. For large-scale connected intelligent devices, mean-filed game with LLM can effectively solve the problem due to the uncertainties of the agent’s behaviour.

B. Use Cases of Collective Intelligence

6G networks with collective intelligence can support many use cases, both in the network and application levels. Autonomous GenAI can largely accelerate network design, planning, implementation, operation, and management, where intent-driven autonomous networks are a typical application. With the capability of planning, LLM can break down a higher level human intents into lower level actionable tasks, to empower an intent-driven autonomous network. For example, an intent of "reduce network energy by 5%" can be converted into actions of tuning transmit power on different sub-carriers, discontinuous reception cycles, channel quality measurements, and so on. This requires the LLM to be adapted with domain knowledge, and grounded to the real network scenario.

Collective intelligence can also empower collaborative robots, vehicles, and devices. For example, connecting LLM on robots allows them to collaboratively serve human instructions. Vehicles connected with LLMs can deliver more reliable and efficient autonomous driving. With multiple on-device LLMs performing perception, planning and control, communication efficiency can be improved as LLM obtains sufficient knowledge grounded to the specific environment. Furthermore, conventional chat-bot tasks such as query, question answering, text generation can be more efficient with collective intelligence from on-device LLMs, which reduces the latency of responses and energy consumption of inferences.

IV. AGI-EMPowered WIRELESS NETWORKS: A Vision FORWARD

The evolution of several AI paradigms is approaching the concept of AGI, in which machines will enjoy a level of intelligence that is equivalent to or surpassing human intelligence. This means that machines will enjoy a broad spectrum of cognition, i.e., will be capable of understanding, learning, applying knowledge, and performing reasoning and inferring in various domains, in an autonomous fashion. It is envisioned that LLMs, with their generalized and generative capabilities, will be the keystone towards the successful deployment of AGI in wireless networks, in which AGI, empowered by LLMs, can be the orchestrator to the efficient planning, design, deployment, configuration, and operation of future wireless networks. In what follows, we put our forward-looking vision of how LLMs, or more generally foundation models, will be integrated into wireless networks, with the aim to realize the true vision of network automation.

A. Task-Agnostic Large Telecom Model

The key driver behind adapting LLMs at the Telecom domain is to exploit multimodal data, generated in wireless networks, to develop Large Telecom Models to enable improved communication and control in wireless networks. In particular, the envisioned large Telecom model is anticipated to act as a general-purpose backbone to the network, with high scalability and flexibility to be deployed on edge devices. In more details, the large Telecom model is foreseen to be designed to perform several general Telecom-oriented tasks, in order to reduce the cost of training multiple AI models to perform specific tasks. Accordingly, the pretrained model can be tailored to fit within a particular downstream task, including modulation, coding, power allocation, beamforming, etc., through fine-tuning with the relevant data (Fig. 3). This represents a stepping stone to implementing LLMs at edge devices, where the fine-tuning process is much less complex and smaller data-set is required.

B. Self-Evolving Networks

The concept of self-evolving networks refers to the networks that are capable of autonomously adapt, change, and evolve with the variations experienced in the network and the surrounding environment. By leveraging AI, these networks will have the capability to optimize their configurations in real-time fashion, according to the users demands and the

Multi-modality Large Language Model

Fig. 3: Large Telecom Model: Towards Improved Edge-Intelligence
network dynamics. This will result in a sustainable performance without the human intervention. With the advent of LLMs, and their capabilities to handle multimodality, the vision of self-evolving networks will go beyond the self-adaptation and self-optimization principles. In specific, we anticipate that LLMs will be leveraged to contribute to the initial steps of designing, planning, deploying, configuring, and operating wireless networks. The use of textual documents from the standards and research reports will allow pretrained LLMs to generate the required software codes and hardware design specification, which will be then go to the deployment stage. In the later stages, LLMs can be employed to identify the optimum configuration of the network according to the initial design, requirements, and users needs. It can be further exploited to build new communication schemes, that are not necessarily compatible with a particular standard, according to a particular network condition, for example, novel waveform to cope with high mobility scenario, modulation and coding design to accommodate multiple high data-rate users, etc. This vision of AGI-empowered wireless networks (Fig. 4) can be achieved with the utilization of the diverse Telecom data modality and the predictability and generative nature of LLMs.

V. CHALLENGES AND OPEN RESEARCH DIRECTIONS

A. Models Compression

Recent LLMs generally consist of at least billions of parameters, e.g., GPT-3 with 175 billion parameters, constraining their deployment in 5G network and beyond, where the number of devices is anticipated to be extremely large. Besides, to meet the requirement of upcoming 6G, and with the increased interest on edge-intelligence, it is essential to develop effective compression mechanisms for Telecom LLMs. This includes pruning, quantization, knowledge distillation, low-rank decomposition and parameter sharing.

B. Data Privacy & Security

With the need to train LLMs using large data-sets representing all user conditions and scenarios, privacy concerns will be raised in which sensitive data pertinent to users need to be shared. Therefore, it is essential to develop privacy-preserving schemes to ensure a maintained data privacy during data collection, processing, and inference. Furthermore, as Telecom LLMs are vulnerable to attacks, it is of paramount importance to ensure robust security measures. This starts with designing a resilient and secure infrastructure that is capable of supporting the language models. Moreover, it is essential to develop secure mechanisms to safeguard training data in rest and in transit. Particular attention should be devoted to Telecom LLM implemented at the edge, where nodes are more compromised to security breaches.

C. Telecom Models Calibration

Models calibration refers to the fine-tuning and parameters adjusting process, in which a task-agnostic language model is tailored to a particular task, in order to enhance the model performance under a particular criteria. Within the Telecom domain, particular concerns is tied to the task-related data, which play an essential role in quantifying the performance of the fine-tuned model. Furthermore, models calibration can be a limiting factor when implementing LLMs in resource constrained networks, where the calibration method should be optimized to fit within the resources budget at the intended devices. This opens a new research direction on developing efficient model calibration methods for edge networks.

D. Compatibility & Adaptivity

Deploying LLM paradigm into current wireless networks might raise several challenges with respect to the compatibility of these models with existing infrastructure. This is due to potential limitations in terms of data, as well as configuration and transmission protocols. Therefore, it is essential to understand to what extent current wireless infrastructure can accommodate LLMs, and what is needed from future wireless networks in order to facilitate the employment of Telecom LLMs.

E. Telecom LLM Architectures

Despite the high number of architectures designed for the LLM, that are capable of handling multimodality, adapting the paradigm of LLMs to the Telecom domain will require sophisticated architectures that are designed and trained from scratch, in order to fit within the Telecom domain. This is stemmed from the fact that Telecom data enjoy unique characteristics, compared to textual data and images/videos. In particular, it is expected that the available LLM architectures are incapable of handling RF data, and integrating it with other data modalities. This call for an urgent need for robust, efficient architectures for LLM in order to enable their implementation in the Telecom domain.

F. Telecom Data

While different types of data modalities are exploited by several existing foundation models, including text, images, and videos, generally domain-specific data is challenging to be incorporated with LLMs, due to the relatively limited sizes of such datasets. Accordingly, such datasets are used for fine-tuning purposes in order to adapt the model into a specific domain. Telecom data is particularly challenging to acquire and to integrate into LLMs due to the involvement of a wide-range of wireless data, that have different nature and characteristics than the earlier mentioned data modalities. In addition to Telecom corpus, this includes RF measurements, geo data, images, videos, and sound signals.

G. Computing Resources & Energy Efficiency

Taking into consideration the extremely large sizes of LLMs and the huge amount of data to be processed, computing and energy resources are two main pillars in developing efficient large telecom models, particular with the inherent limitation on energy resources in wireless networks. Several techniques can be adopted to optimize the model efficiency in order to reduce the computational requirements, including model compression.
and quantization, and knowledge distillation. By exploiting the network architecture, various schemes can be adopted for optimizing the needed resources for models training as well as for inference. This includes, tasks offloading, dynamic resource allocation, and optimized communication protocols.

VI. CONCLUSION

In this article, we explored how LLMs can be an essential tool in designing, configuring, and operating future wireless networks. In particular, we identified the key opportunities, with respect to sensing and communication, that can be acquired when employing LLMs in wireless networks, and we overviewed the role of wireless networks in enabling machines to communicate using LLMs. Moreover, we laid down the foundation for the development of the AGI-empowered wireless networks through LLMs, which paves the way to the successful implementation of self-evolving networks. We finally opened new horizons for future research directions that need to be further investigated to realize the true vision of Telecom LLM.

REFERENCES


BIographies

Lina Bariah (Lina.Bariah@tii.ae) is a Senior Researcher at Technology Innovation Institute, UAE.
Qiyang Zhao (Qiyang.Zhao@tii.ae) is a Lead Researcher at Technology Innovation Institute, UAE.
Hang Zou (Hang.Zou@tii.ae) is a Researcher at Technology Innovation Institute, UAE.
Yu Tian (Yu.Tian@tii.ae) is a Researcher at Technology Innovation Institute, UAE.
Faouzi Bader (Carlos-Faouzi.Bader@tii.ae) is the Director of Telecom Unit at Technology Innovation Institute.
Merouane Debbah (Merouane.Debbah@ku.ac.ae) is a Professor at Khalifa University, UAE.