RÉPUBLIQUE ALGERIENNE DÉMOCRATIQUE ET POPULAIRE
MINISTÈRE DE L'E NSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE SCIENTIFIQUE
UNIVERSITÉ Saad Dahleb, Blida 1

— Institut Aéronautique et Études spatiales IAES —

Mémoire

Présentée pour obtenir le diplôme de

MASTER 02 LMD

SPÉCIALITÉ : TÉLÉCOMMUNICATION

LLMs-TBSA: Large Language Models for Troubleshooting of Base Stations Anomalies

Par

Abir Azzouz

Soutenue le 04/07/2024, devant le jury composé de:

- Dr. Krim Mohamed, Président.
- Dr. Benaissa Rachid, Examinateur.
- Dr. Asma Iman Kouachi, Encadreur.

Promotion 2023-2024

Acknowledgement

In the name of Allah, the Most Gracious, the Most Merciful.

First, I am faithfully and deeply grateful to Allah, the Almighty, for granting me the opportunity, strength, and perseverance to complete this dissertation successfully.

I would like to express my heartfelt gratitude to my parents, **Mohamed Azzouz and Noura Djebbar**, for their constant encouragement and for believing in me throughout this journey. Their support has been an invaluable source of motivation.

I am deeply indebted to my supervisor, Professor **Asma Iman Kouachi**, for her exceptional guidance, patience, and encouragement during this research. Her expertise and constructive feedback have been instrumental in shaping this dissertation.

I am also grateful to Professor **Sofiane Tahraoui** for his valuable assistance and support throughout these years. The institution is truly fortunate to have a scholar of his caliber.

I would like to thank the committee members:**Dr. Krim Mohamed and Dr. Benaissa Rachid** for their acceptance to discuss my work.

Finally, I would like to extend my heartfelt thanks to my brother **Ayoub**, my brother **Oussama**, and my sister **Roufaida**. I am truly blessed to have you in my life and your love and encouragement have been a constant source of strength.

الإهداء

الحمد لله الذي ما نجحنا و ما علونا و لا تفوقنا إلا برضاه

الحمد لله الذي ما إجتزنا دربا و لا تخطينا صعبا إلا بفضله

"وَآخِرُ دَعْوَاهُمْ أَنِ الْحَمْدُ لِلَّهِ رَبِّ الْعَالَمِينَ "

من قال أنا لها "نالها " وأنا لها إن أبت رغما عنها أتيت بها

و في هذه اللحظة الاكثر فخرا أهدي عملي هذا الى من أحمل إسمه بكل إفتخار,الى من كلله الله بالهيبة و الوقار ،و علمني أن النجاح لا ياتي إلا بالصبر و الإصرار ،الى فخري و إعتزازي -والدي الغالي-

إلى قدوتي الأولى،وشمعتي في الليالي المظلمة، إلى الإنسانة العظيمة التي طالما تمنت أن تقر عينها برؤيتي في يوم كهذا ،الى سر قوتي و نجاحي -والدتي الحبيبة-

أشكر ملهمتي الأولى أستاذتي العزيزة ،السيدة- أسماء إيمان كواشي -التي كانت معي في كل خطوة خطوتها و لها الفضل الكبير في إتمام هذا البحث

إلى ضلعي الثابت وأمان أيامي ،إلى الداعم الاول الذي هون علي كل أزمة و مشقة ،إلى الصديق الوفي الذي كلما إحتجت إليه كان حاضرا بلا تردد- أخي العزيز- شكرا لوجودك الحنون ، أدامك الله لنا سندا لا يميل

إلى من شددت عضدي بهم فكانو لي ينابيع أرتوي منها، إلى خيرة ايامي و صفوتها ،إلى قرة عيني -إخوتي-

إلى الداعم و المشجع لهذا الانجاز ،إلى رفيق الدرب ادامك الله عمرا مديدا

إلى صديقة المواقف لا السنين ،شريكة الدرب و الطموح البعيد - صديقتي العزيزة-

لكل من كان عوناً و سندًا في هذه الطريق

أهديكم هذا الإنجاز وثمرة نجاحي التي لطالما تمنيته،

و ها أنا أتممت أول ثمراته

راجية من الله تعالى أن ينفعني بما علمني وأن يعلمني ما أجهل و يجعله حجة لي لا علي

Abstract

The telecommunications sector is a vast global industry, fraught with numerous anomalies and challenges that companies must manage. A critical component of the mobile network infrastructure is the Base Station (BS). In Algeria, mobile operators such as Mobilis, Ooredoo, Djezzy, and Telecom Algeria continue to address BS anomalies using traditional methods, which can be inefficient and time-consuming.Large language models (LLMs) have recently demonstrated significant promise across various domains due to their advanced capabilities in understanding and generating human-like text. This dissertation investigates the application of LLMs within the telecommunications industry to resolve BS anomalies intelligently.In our research, we implemented three LLMs: Llama, GPT-2, and Mistral—to generate solutions for BS anomalies. Our results demonstrate that the Mistral model outperforms the others in effectively resolving BS anomalies, highlighting its potential to revolutionize how telecommunications issues are addressed.

Keywords: Telecom, Large Language Models (LLMs), Base Stations Anomalies (BSA), Mistral, Llama, GPT 2.

Résumé

Le secteur des télécommunications est une vaste industrie mondiale, confrontée à de nombreuses anomalies et défis que les entreprises doivent gérer. Un composant essentiel de l'infrastructure des réseaux mobiles est la station de base.En Algérie, les opérateurs mobiles tels que Mobilis, Ooredoo, Djezzy et Algérie Télécom continuent de traiter les anomalies des stations de base en utilisant des méthodes traditionnelles, souvent inefficaces et chronophages.Les grands modèles de langage se sont récemment révélés très prometteurs dans divers domaines en raison de leurs capacités avancées de compréhension et de génération de textes semblables à ceux des humains. Cette thèse étudie l'application des LLM dans l'industrie des télécommunications pour résoudre intelligemment les anomalies des stations de base.Dans notre recherche, nous avons implémenté trois modèles : Llama, GPT-2 et Mistral, pour générer des solutions aux anomalies des stations de base. Nos résultats démontrent que le modèle Mistral est plus performant que les autres dans la résolution efficace des anomalies des stations de base, soulignant son potentiel à révolutionner la façon dont les problèmes de télécommunications sont traités.

Mots clés: télécommunications, grand modèle de langage, anomalies des stations de base, Mistral, Llama, GPT-2.

ملخص

يعد قطاع الاتصالات صناعة عالمية واسعة النطاق، مليئة بالعديد من المشاكل والتحديات التي يجب على الشركات إدارتها. أحد المكونات المهمة للبنية التحتية لشبكة الهاتف المحمول هو المحطة الأساسية (BS).في الجزائر، يواصل مشغلو الهواتف المحمولة مثل موبيليس، وأوريدو وجيزي واتصالات الجزائر معالجة المشاكل في نظام BS باستخدام الطرق التقليدية، والتي قد تكون غير فعالة وتستغرق وقتًا طويلاً.أثبتت النماذج اللغوية الكبيرة (LLMs) مؤخرًا أنها واعدة بشكل كبير في مختلف المجالات نظرًا لقدراتها المتقدمة في فهم وتوليد نصوص شبيهة بالنصوص البشرية. تدرس هذه الأطروحة تطبيق النماذج اللغوية الكبيرة في صناعة الاتصالات السلكية واللاسلكية لحل المشاكل في نظام BS بذكاء.في بحثنا، قمنا بتطبيق ثلاثة نماذج : LLAMA و GPT-2 و Mistral لتوليد حلول لمشاكل BS. تُظهر نتائجنا أن نموذج Mistral يتفوق على النماذج الأخرى في حل المشاكل في نظام BS بكفاءة، مما يسلط الضوء على قدرته على إحداث ثورة في كيفية معالجة مشاكل الاتصالات السلكية واللاسلكية.

الكلمات المفتاحية: الاتصالات، نماذج اللغات الكبيرة ،مشاكل المحطات الأساسية،GPT 2، Llama، Mistral.

Contents

List of Figures

List of Tables

Acronyms

- **1G** : First Generation
- **2G** : Second Generation
- **3G** : Third Generation
- **3GPP** : The Third Generation Partnership Project
- **4G** : Fourth Generation
- **5G** : The Fifth-generation
- **5GC** : 5G Core
- **6G** : The Sixth-generation
- **AC** : Alternating Current
- **AI** : Artificial Intelligence
- **AMPS** : Advance Mobile Phone Service
- **APE** : Absolute Positional Embedding
- **AUC** : Authentication Center
- **BBU** : Base Band Unit
- **BERT** : Bidirectional Encoder Representation from Transformer
- **BSC** : Base Station Controller
- **BSS** : Base Station Subsystem
- **BTS** : Base Transceiver Station
- **CDMA** : Code Division Multiple Access
- **CN** : The Core Network

- **DC** : Direct Current
- **DMR** : Digital Mobile Radio
- **E-UTRAN** : The Evolved Universal Terrestrial Radio Access
- **EPC** : Evolved Packet Core
- **ETSI** : The European Telecommunications Standards Institute
- **FDMA** : Frequency Division Multiple Access
- **GenAI** : Generative Artificial Intelligence
- **GNMT** : Google Neural Machine Translation
- **GPS** : Global Positioning System
- **GPT** : Generative Pre-trained Transformer
- **GSM** : Global System for Mobile Communications
- **HAP** : High-Altitude Platforms
- **HLR** : Home Location Register
- **HTS** : High Throughput Satellites
- **IoT** : Internet of Thing
- **LAP** : Low-Altitude Platforms
- **LGMs** : Large GenAI Models
- **LLMs** : Large Language Models
- **LLaMA** : Large Language Model Meta AI
- **LTE** : Long-Term Evolution
- **MC** : Mobile Core
- **MCN** : Mobile Communication Network
- **MSC** : Mobile Switching Center
- **MS** : Mobile Station
- **MU** : Mobile Unit
- **MTSO** : Mobile Telephone Switching Office
- **NLP** : Natural Language Processing

- **NTN** : Nonterrestrial Networks
- **NSS** : Network and Switching Subsystem
- **PaLM** : Pathways Language Model
- **QoS** : Quality of Service
- **RAGs** : Integration of Retrieval-Augmented Generations
- **RAN** : Radio Access Network
- **RCU** : Remote Control Unit
- **RET** : Remote Electrical Tilt
- **RF** : Radio Frequency
- **RFU** : Radio Frequency Unit
- **RNS** : The Radio Network Subsystem
- **RNNLM** : Recurrent Neural Network Language Model
- **RRUs** : Remote Radio Units
- **SFT** : Supervised Fine-Tuning
- **SIM** : Subscriber Identity Module
- **SLMs** : Small Language Models
- **T5** : Text-to-Text Transfer Transformer
- **TDMA** : Time Division Multiple Access
- **TT ID** : Trouble Ticket Identifier
- **UAVs** : Unmanned Aerial Vehicles
- **UE** : User Equipment
- **UMTS** : Universal Mobile Telecommunications System
- **UTRAN** : Universal Terrestrial Radio Access
- **VLR** : Visitor Location Register

Introduction

1. Context

Within the telecommunications industry, mobile networks serve as central infrastructures connecting individuals, businesses, and communities on a global scale.

The rapid evolution of mobile networks has been driven by technological advancements and increasing demands for enhanced communication capabilities. Progressing from 2G to 3G, 4G, and now anticipating 5G and 6G networks, these advancements have significantly improved connectivity, offering faster, more reliable, and more efficient communication channels.

Large Language Models (LLMs) represent a breakthrough in Artificial Intelligence (AI), capable of understanding and generating natural language text, performing various language tasks, and demonstrating profound language comprehension and generation abilities [11]. These models have revolutionized fields such as natural language processing, text generation, chatbots, content summarization, and translation, fundamentally altering human-machine interactions.

Integrating LLMs into the telecommunications sector holds transformative potential, enhancing operational efficiency and service delivery. LLMs enable applications such as customer service chatbots and automated troubleshooting, promising advancements in communication technologies.

2. Problem Description

Base Station (BS) anomalies encompass irregularities within the critical components of wireless communication networks. These anomalies, which include hardware failures, software glitches, and environmental factors, significantly impact network performance, reliability, and service quality [17, 18].

Telecommunication networks encounter various anomalies, manifesting as dropped calls, slow data speeds, coverage gaps, and other issues that degrade user experience. Identifying and resolving these anomalies efficiently is challenging due to the vast amount of data generated by modern telecom networks. Leveraging LLMs offers a transformative approach to address these challenges by analyzing large datasets, logs, and customer interactions [19, 20].

Traditional methods for handling BS anomalies often rely on manual troubleshooting and simplistic detection mechanisms, which are inadequate for the complexities of modern telecommunications. Therefore, there is a pressing need for advanced, automated solutions to swiftly diagnose and resolve BS anomalies, ensuring optimal network performance and uninterrupted service delivery.

3. Contribution

This dissertation makes a groundbreaking contribution to the telecommunications industry by integrating LLMs to address base station (BS) anomalies.

We have selected the three most widely used LLMs in the telecommunications domain: GPT-2, Llama, and Mistral. These models are employed to generate solutions for BS anomalies.

Here are some advantages of using LLMs for generating solutions to base station (BS) anomalies:

(a) **Automated Resolution:** LLMs can automatically resolve anomalies in real time, reducing the need for manual intervention and enabling faster response times.

(b) **Accuracy:** Advanced LLMs can provide highly accurate anomaly diagnosis by learning from vast datasets, leading to precise identification and resolution of issues.

(c) **Cost Efficiency:** Automating the resolution process reduces operational costs by minimizing the need for human intervention and reducing downtime.

(d) **Integration with Existing Systems:** LLMs can be integrated with existing network management systems, allowing for seamless implementation and utilization of their capabilities without requiring significant changes to the current infrastructure.

BTW, This contribution is under review for an international conference.

3. Dissertation Structure

This dissertation is structured as follows:

- Chapter 1: This chapter provides essential background information on mobile networks.
- Chapter 2: Explores the fundamentals of LLMs and reviews existing literature and studies that integrate LLMs within the telecommunications field. The sections included in this chapter are:
- Chapter 3: Focuses on our proposed solutions, evaluation metrics for selected models, and presents findings from evaluations. It concludes with discussions and outlines future research directions.

Chapter 1

Background on Mobile Network

Contents

1.1 Introduction

Naturally, humans are social and interact to form relationships, share ideas, and exchange information.

Communication networks include various methods and technologies that people use to communicate and interact with each other. Throughout history, people have used different ways to communicate with each other, such as fax, telegraph, telefax, radio systems, and mobile phones.

Mobile communication networks have become the most common way that people communicate with each other because of their flexibility in being used anywhere and anytime [21].

In this chapter, we will shed light on:

- In Section 1.2, we provide a basic definition of mobile networks.
- In Section 1.3, we discuss the concept of generations of mobile networks, followed by their architecture.
- In Section 1.4, we describe the different applications offered by mobile networks.
- In section 1.5, we explain the main equipment of the base station.
- Finally, we conclude this chapter in Section 1.6.

1.2 Definition of Mobile Network

- Also known as Mobile Communication Networks (MCNs), they are a category of telecommunications networks [21].
- Mobile communication systems have revolutionized how people communicate by combining communication capabilities with mobility [5].
- The telecommunications industry has grown through various generations, including the first generation (1G), the second generation (2G), the third generation (3G), the fourth generation (4G), the fifth generation (5G), and the sixth generation (6G). Each generation has its capabilities and techniques, distinguishing it from the previous[22].
- Mobile networks began as two-way radio systems and evolved into 4G networks capable of transmitting high-quality video and voice calls [21].

1.3 Mobile Network Evolution

This section explains the main idea of mobile network generations followed by their architectures:

1.3.1 The First Generation

- **Definition of First Generation (1G):**

  In the 1970s, Amos Edward Joel developed the 1^{st} generation (1G) [21]. 1G is the first wireless technology, commonly known as cell phones. It was developed in the 1980s [23].

The 1G, known as Advance Mobile Phone Service (AMPS) technology, was an analog system. AMPS uses frequency modulation and Frequency Division Multiple Access (FDMA). It had a channel capacity of 30 KHz and operated in the frequency band of 824-894 MHz [22].

- **Architecture of 1G:**

  As you can see in Fig.1.1 and according to[24] [1], the key components of the 1G cellular radio system architecture are:

  - Mobile Unit (MU): the MU, such as an early cell phone, is the device subscribers use to communicate within the cellular network.
  - Base Transceiver Station (BTS): each cell has a BTS, a tower that transmits signals to and receives signals from mobile units within its coverage area.
  - Mobile Telephone Switching Office (MTSO): BTSs link to an MTSO, which then links to local exchanges responsible for routing calls to their destination. The links between the BTSs and the MTSO can be microwave or wireline.

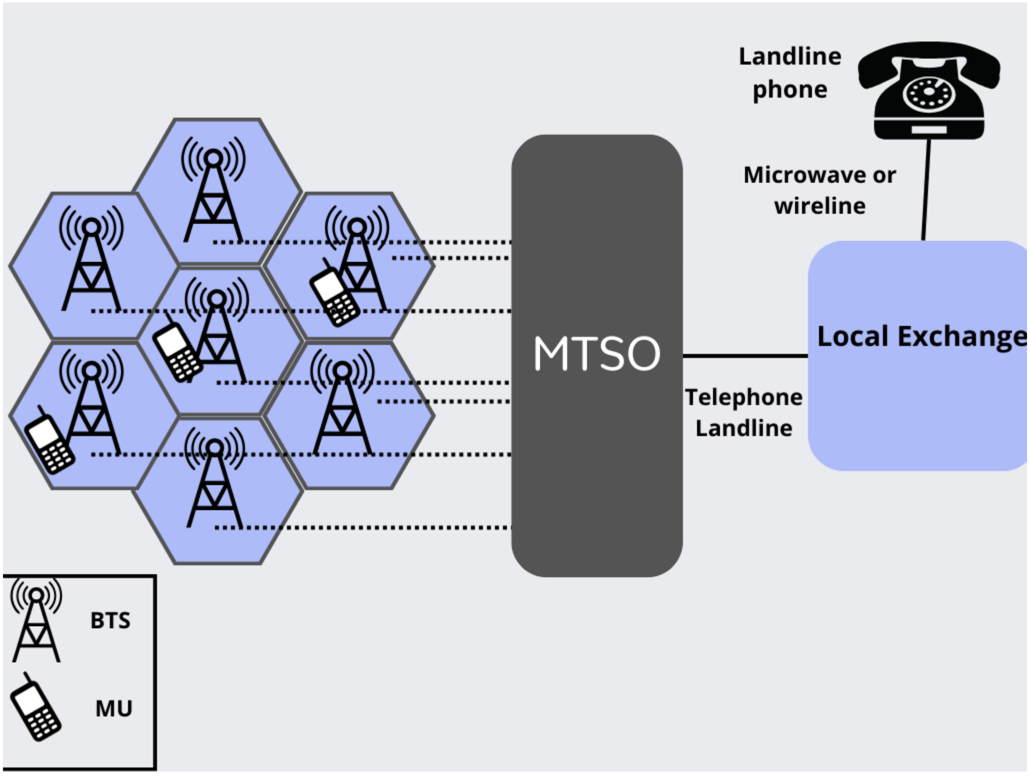

Figure 1.1: 1G architecture [1, 2].

1.3.2 The Second Generation

- **Definition of Second Generation (2G):**

  The 2nd generation (2G) mobile communication system, launched commercially in Finland in 1991, introduced digital technology [22].

  The 2G utilized the Global System for Mobile Communications (GSM) technology[21]. This system is still used mainly in various regions of the world.

  The GSM is the international standard for mobile communication, developed by the European Telecommunications Standards Institute (ETSI) in the 1980s. It allows users to make phone calls and send text messages.[25]

  This generation primarily provides voice and data (message) services. It employs two digital modulation schemes: (1) Time Division Multiple Access (TDMA) and (2) Code Division Multiple Access (CDMA)[22].

- **Architecture of 2G:**

  This architecture is also called the GSM architecture. According to [3, 26, 6], the GSM architecture consists of three main subsystems as shown in Figure 1.2, which are:

  1. Mobile Station (MS): The MS is the user's hardware, like a mobile phone or tablet, responsible for communicating with the network and providing services to the user. The two most essential elements of the MS are the hardware and the Subscriber Identity Module (SIM) card.
  2. Base Station Subsystem (BSS): it provides signal coverage to mobile stations and is composed of two components:
     - Base Transceiver Station (BTS): The BTS is the radio interface between the MS and the network and is responsible for the transmission and reception of radio signals.
     - Base Station Controller (BSC): It oversees the management of multiple BTSs. The BSC represents the link connecting mobile devices to the Mobile Switching Center (MSC).
  3. Network and Switching Subsystem (NSS): It represents the core network of the GSM system, its main role is switching calls between base stations. It consists of the following components:
     - The Mobile Switching Center (MSC): MSC is a key component of NSS, which provides call setup, call routing, and call switching.
     - Home Location Register (HLR): The HLR is a central database that contains information for users within a specific area. HLR is used to locate users and provide services that are specific to each subscriber.
     - Visitor Location Register (VLR): It stores subscriber information for users roaming in a specific area.

- Authentication Center (AUC): The AUC provides verification and encryption to ensure the user's identity and keep their calls private. The verification center is a secure file with each user's secret key stored in their SIM card.

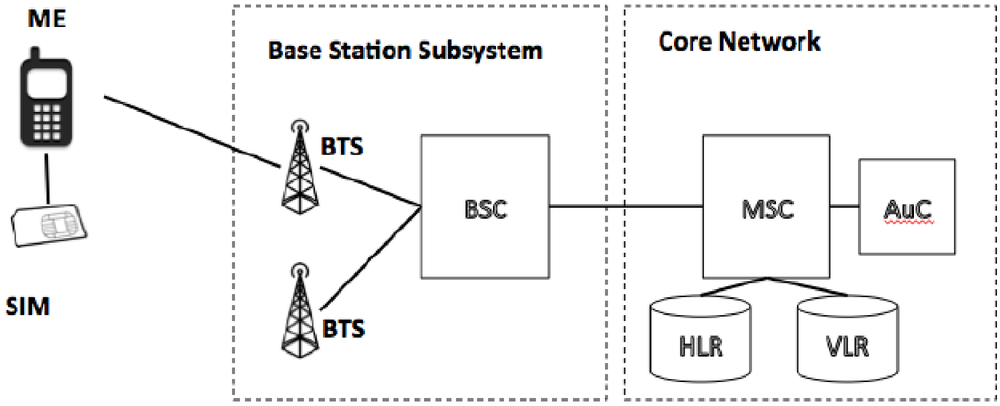

Figure 1.2: 2G architecture [3].

- **Comparison between 1G and 2G:**

  The key points on the differences between the 1G and 2G are:

  1. **Voice Quality**: 1G networks had poor voice quality due to analog signals which are continuous signals that represent information in a range of values, while 2G networks offered improved voice quality with digital signals which are discrete signals representing information in binary form [27].
  2. **Security**: 1G networks had no security, making calls susceptible to eavesdropping. 2G networks provided better security measures[23].
  3. **Connectivity**: 1G analog signals could maintain connectivity over longer distances but with lower call quality. In contrast, 2G digital signals depended more on location and proximity to cell towers [23].

1.3.3 The Third Generation

- **Definition of Third Generation (3G):**

  In 2007, 3rd generation (3G) mobile technology enabled users to access audio, graphics, and video applications. This advancement in 3G technology allowed users to watch videos and make video calls. The 3G is also called the Universal Mobile Telecommunications System (UMTS) technology [21].

  UMTS is a mobile communication technology that was introduced in 1999, based on the GSM standards, It is designed to offer faster data speeds in download and support a wider range of services, including internet access [4].

The 3G mobile system is appropriate with various cellular standards, including CDMA, GSM, and TDMA, under a unified framework [22].

The transition from 2G to 3G involves gradually improving the GSM network and services to provide more functionality, options, and value. This evolution in mobile technology aims to achieve high-speed data transmission, fast data rates, and good Quality of Service (QoS) in the 3rd generation mobile communication system [22].

- **Architecture of 3G:**

  This architecture is also called the UMTS architecture. According to [4, 6], the 3G UMTS wireless communication system saw significant changes from the previous generation. Its top-level network architecture can be divided into three main elements, as shown in the figure below 1.3:

  1. The User Equipment (UE): The UE in the 3G UMTS system, equivalent to the mobile station in GSM, refers to the device previously known as a mobile or cellphone. The term UE was adopted because of its increased functionality and capabilities.
  2. The Radio Network Subsystem (RNS):Is also named UMTS Radio Access Network (UTRAN), equivalent to BSS in GSM.
  3. The Core Network (CN): Is responsible for all central processing and management functions in the 3G UMTS system, similar to the role of the NSs in GSM.

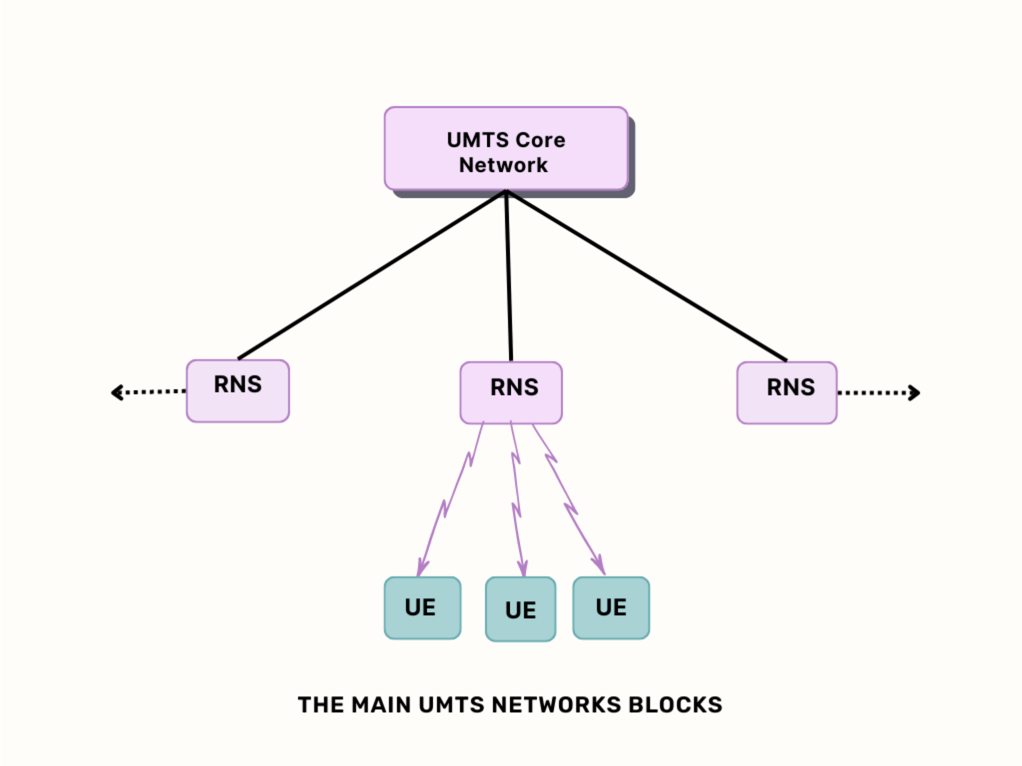

Figure 1.3: 3G architecture [4].

1.3.4 The Fourth Generation

- **Definition of Fourth Generation (4G):**

  The 4th generation (4G) mobile communication services began in 2010, and the huge growth in mobile subscriptions led researchers and industries to focus on the next generation of mobile wireless technology [22, 5].

  This generation has been introduced to provide high transmission rates while ensuring QoS features [21], where the primary goal of 4G technology is to deliver high-speed, high-quality, high-capacity, and cost-effective services [22].

- **Architecture of 4G:**

  This architecture is also called Long Term Evolution (LTE) architecture. The 3rd Generation Partnership Project (3GPP) was created in 1998 and brings together seven telecommunications standard development organizations, known as 'Organizational Partners'[28, 17]. The 3GPP creates detailed specifications for cellular technologies, including radio access, core networks, and services.

  3GPP started exploring LTE technology in 2004, which provides several advantages over other wireless technologies, it is designed to provide faster data speeds compared to previous systems[29].

LTE has a simplified architecture developed by 3GPP called System Architecture Evolution (SAE) [6]. Figure 1.4 shows the LTE architecture which contains:

- Evolved Universal Terrestrial Radio Access Network(E-UTRAN): E-UTRAN is the 4G radio access network, composed of BSs that communicate with UEs and manage tasks such as radio resources and security [30].

  These BSs are interconnected via the X2 interface for communication between them, where the X2 represents the interface that connects two BSs directly and is responsible for handling services like mobility management and error reporting between them[31]. While the S1 interface links BSs to EPC, enabling the exchange of data and control signals [30].

- Evolved Packet Core (EPC): The most important part of the SAE architecture is the EPC, which offers the following functions: connecting to other networks and managing service quality.

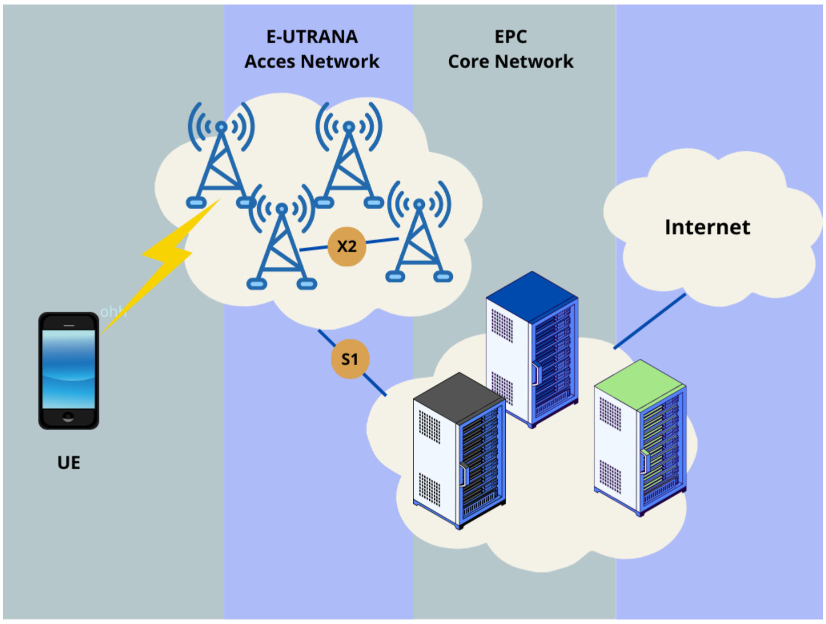

Figure 1.4: 4G architecture [5, 6]

1.3.5 The Fifth Generation

- **Definition of Fifth Generation (5G):**

  The 5G is also known as the 5[th] generation of mobile networks or wireless systems, represents the next significant advancement in mobile communication standards that exceed those of 4G [23].

The advancements in telecommunications from 1G to 5G bring many upgrades in telecommunications. 5G is now available at affordable rates, offering high peak speeds and greater reliability compared to its predecessors [23].

5G technology features advanced concepts such as the Internet of Things (IoTs), low bit cost, and high energy efficiency. These innovations enable high-quality connections, efficient data transfer, and reduced power consumption [21].

- **Architecture of 5G :**

  According to [7] and the figure 1.5, the mobile cellular network is made up of two main parts :

  1. Radio Access Network (RAN): The RAN manages radio resources and ensures quality service for all users through a network of base stations. In 4G, these base stations are called Evolved Node B (eNB), and in 5G, they are called gNB (with "g" meaning next Generation) in 5G.

     The mobile cellular network offers wireless connectivity to devices, including UE, which historically included mobile phones and tablets, but now encompasses a wide range of devices such as cars, drones, industrial machines, robots, home appliances, and medical devices, even when they are in motion.

  2. The Mobile Core (MC): The mobile core network is a central part of the overall mobile network architecture that performs critical functions such as:

     - Authenticates devices before connecting them to the network.
     - Provides Internet connectivity for data and voice services.
     - Ensures connectivity meets QoS requirements.
     - Tracks user mobility for uninterrupted service.
     - Monitors subscriber usage for billing and charging.

     In 4G, this is called EPC; in 5G, it is called the 5G Core (5GC). The Mobile Core bridges the RAN and the Internet, often serving a metropolitan area.

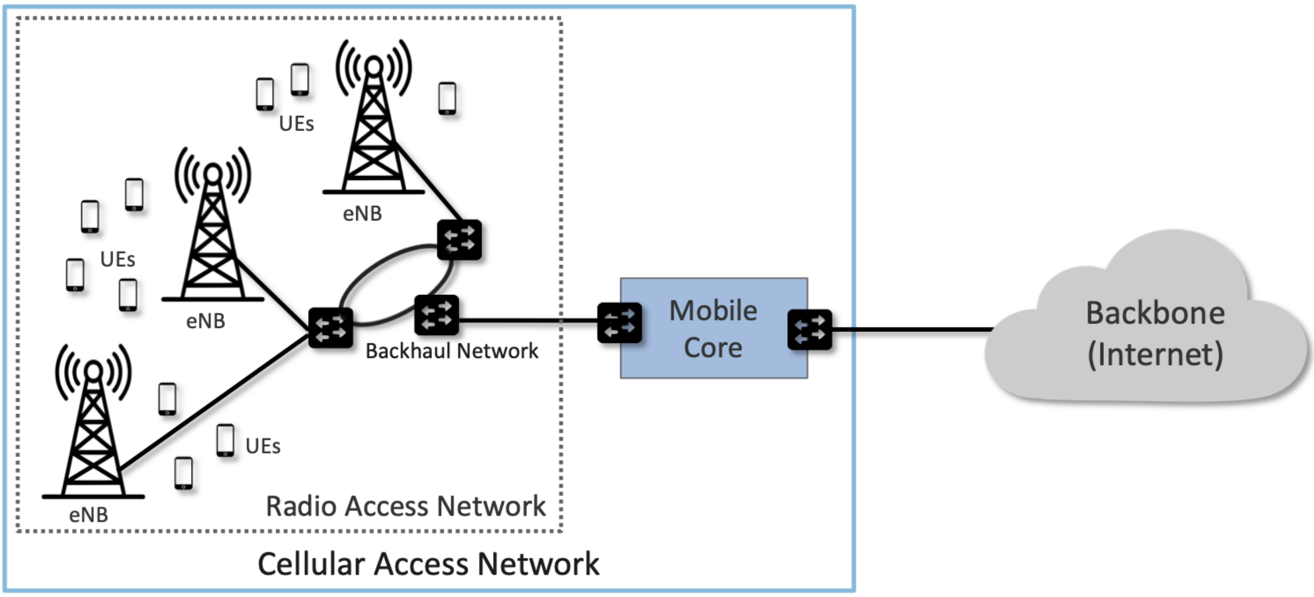

Figure 1.5: 5G Architecture [7].

1.3.6 The Sixth Generation

- **Definition of Sixth Generation (6G):**

  The 6G is also known as the 6th generation of mobile networks. According to [22], the upcoming 6G mobile system, aiming for worldwide coverage, will combine the existing 5G mobile system with satellite networks.

  These satellites serve various purposes: telecommunication for calls and data, the Internet and broadcasting, Earth imaging for weather and environmental data, and navigation for the Global Positioning System (GPS). These satellite systems were developed by four different countries: the USA for GPS, China for COMPASS, the EU for Galileo, and Russia for GLONASS[22].

- **Architecture of 6G :**

  According to [32], the aim of designing the next generation network is to improve the coverage of the communication. Current networks mostly rely on old terrestrial cellular infrastructure, so the plan for 6G is to integrate nonterrestrial networks(NTN)to ensure full wireless connectivity.

  NTN, which includes satellites, drones, and others, offers a promising solution to the challenges faced by the old regular terrestrial networks. That helps expand coverage and capacity of wireless communication [33].

The components of the 6G Network Satellite Air Terrestrial Networks (SATN) architecture, as shown in Figure 1.6 and [8], include:

- Space-Network:High-speed satellites in space, known as High Throughput Satellites (HTS), provide internet like on land, but regular satellites are far away, so the Internet is slow and not good for mobile phones. A new idea with satellites closer to Earth could make the Internet faster than fiber-optic cables. This system might use radio waves and lasers to make the Internet superfast.
- Aerial-Network: Aerial networks come in two types: high-altitude platforms (HAP) high up in the sky and low-altitude platforms (LAP) closer to the ground. HAP covers more and lasts longer, while LAP with Unmanned Aerial Vehicles (UAVs) is quick to deploy and adaptable. UAV networks are handy in emergencies, and new path optimization techniques save energy.
- Ground sea network: BS and UAV work together to help users on the ground connect to the network without causing interference [34]. On the other hand, there are underwater wireless networks, which are tough, with poor coverage, signal weakening, and equipment damage. There are many problems to solve before the 6G technology can be fully developed and ready for worldwide use [32].

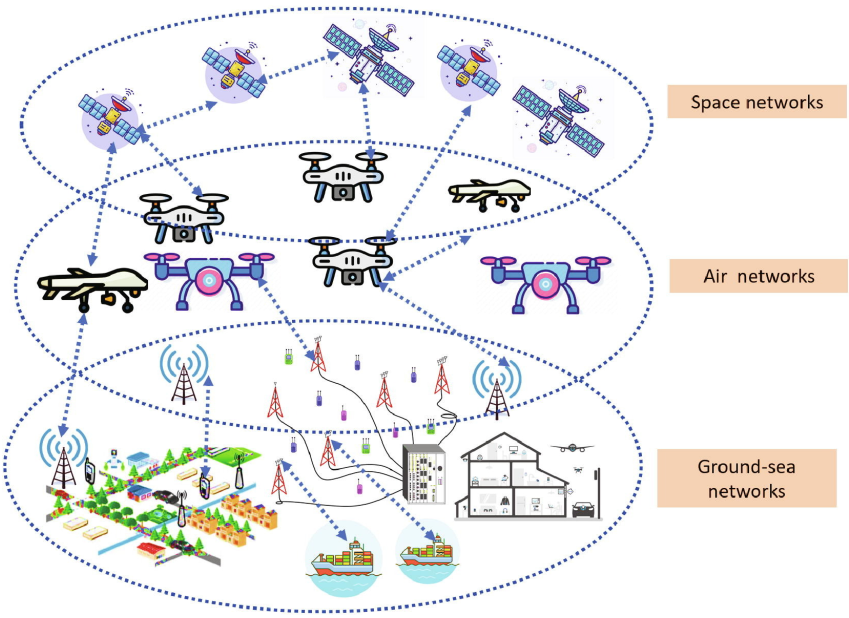

Figure 1.6: 6G Architecture [8].

1.4 Mobile Network Applications

According to [21], mobile networks can be used for many services, such as:

- Medical management (accessing test results, medical records).
- Financial management (paying bills, transferring money).
- Educational support (study courses, learning languages).
- Social connectivity (keep up-to-date with family news and meet new people).
- Business interactions (holding meetings, collaborating with partners).
- Knowledge-sharing and networking opportunities.

1.5 Base Station Components

In this section, we outline the key components of BS that will be referenced in subsequent chapters.

According to [9] BSs are crucial for setting up network services for users because they transmit signals, helping to create a full network system.

The components of a 4G LTE base station, as shown in general in figure 1.8 it is in detail are as follows:

- The outdoor equipment: the outdoor equipment of a BS refers to the components installed outside including :
  - The Antenna system: typically located at the top of the BS, is the primary component responsible for transmitting and receiving radio signals and is connected to The Remote Radio Units (RRU) for signal processing and management, shown in the figure1.7. A base station usually has three antennas, each covering a 120-degree area, which together provide complete 360-degree wireless coverage around the base station [35].
  - Remote Electrical Tilt (RET): Antennas often have RET capabilities, enabling network engineers to adjust the tilt angle remotely for better signal optimization. This feature is managed by a Remote Control Unit (RCU) that communicates with the RRU [36].
  - Feeders: they are coaxial cables that transmit Radio Frequency (RF) signals from the RRU to the antenna. As illustrated in Fig.1.7 [36].

- Remote Radio Unit (RRU): Also known as Radio Frequency Unit(RFU)[35]. According to [36] The RRU links to the antenna to transmit RF signals, with separate connections for sending and receiving. The RRU converts RF signals to digital data for transmission to the Base Band Unit (BBU); we explain BBU later; and vice versa, involving amplification and filtering for signal quality. It is placed close to the antenna, often on the tower, to minimize signal loss and enhance efficiency.

  The connection to the BBU is established via a bidirectional fiber optic connection, typically using the Common Public Radio Interface (CPRI) for fast, reliable, and low-latency communication.

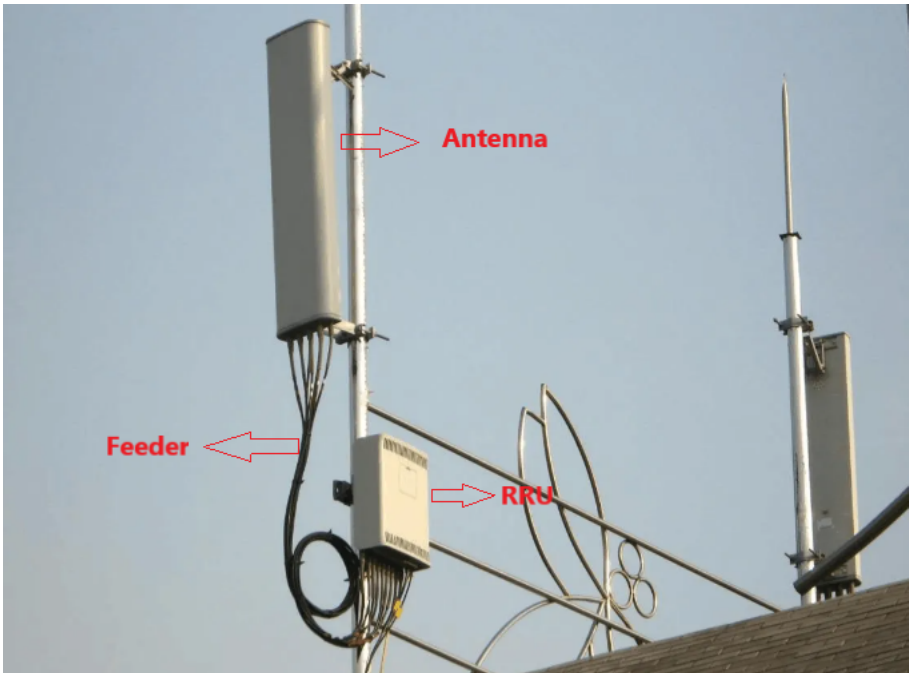

Figure 1.7: Antenna and RRU System [9].

- The indoor equipment: The indoor equipment of a BS refers to all the components installed inside a building or an equipment room or cabinet, including the BBU, power supply system, and cooling systems [9].
  - The machine room: The machine room's main job is to protect and house vital equipment. It also has strict access control systems for security. For equipment work, a power source is necessary, initially, AC (Alternating Current) power is obtained from the city's power grid, which is then converted into DC (Direct Current) power [35]. Air conditioning is necessary to maintain optimal temperatures in the machine room due to the heating introduced while the equipment is working [35]. the machine room contains the following components:

- BBU (Baseband Unit): The BBU is located in the machine room[9]. According to [36], the BBU is the brain of the base station. It processes signals, manages the base station's operations, and connects the BS to the core network by handling signal processing, control functions, and system maintenance.
- DC power supply system: DC power supply system contains essential components like a power system, backup batteries in case of power failure, transmission equipment, and air conditioning [9]. DC is connected to the RRU and the BBU with DC cables.
- Transmission equipment: Is needed to connect the base station to other base stations in the network [35].

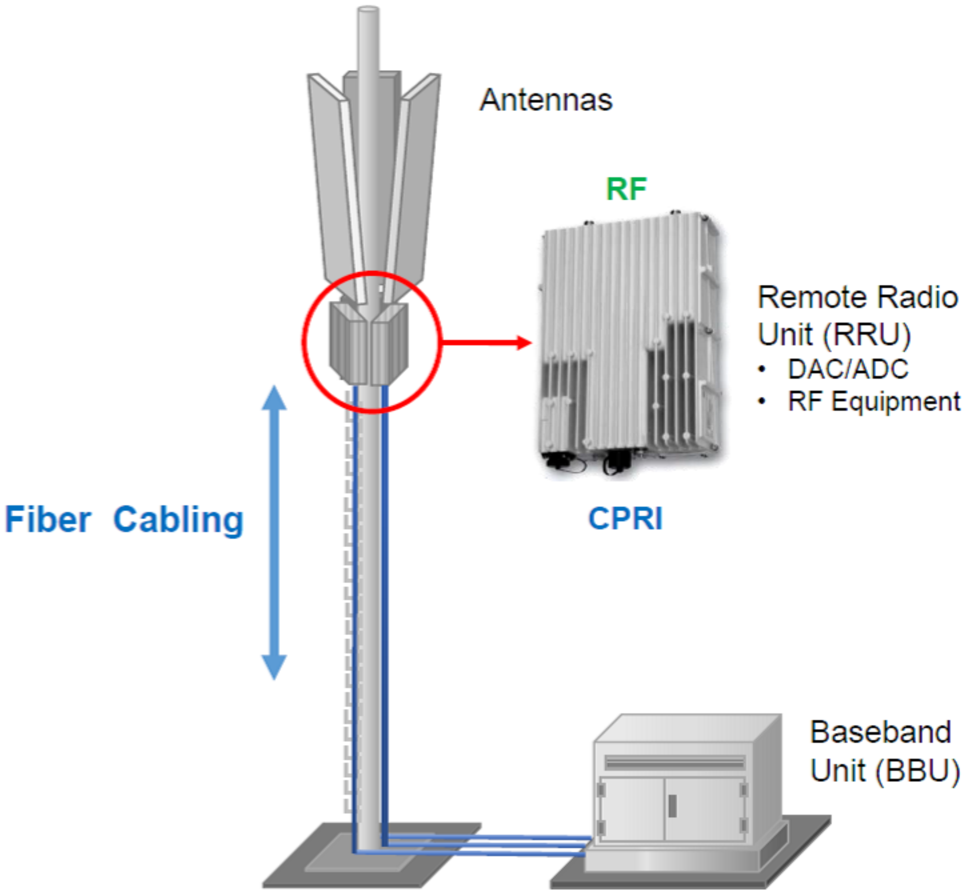

Figure 1.8: Base Station equipments [10].

1.6 Conclusion

In conclusion, mobile phones have become crucial in our daily lives, evolving through various generations of wireless technology.

Mobile technology has evolved significantly over the years, with each generation building upon the previous one. The rapid advancement of wireless networks is driven by the increasing demands and expectations of users, as well as the evolution of Internet traffic and applications.

Chapter 2

Background on Large Language Models (LLMs)

Contents

2.1 Introduction

Language is an important human ability that begins in childhood and develops over time. However, machines cannot understand or communicate in human language without Artificial Intelligence (AI) algorithms. They need AI to understand and talk to people.

Large Language Models (LLMs) are advanced computer programs trained on vast amounts of text data. They can understand and generate human-like language, which allows them to perform tasks like answering questions, translating languages, generating text, etc [37, 38, 39].

Telecommunications rapidly connect people globally, drive economic growth via digital services, and foster innovation with technologies like 5G and IoT. Essential for emergency responses and enhancing social networks, it is indispensable in modern life.

In the year 2024, there are many papers which they focus on using LLMs in the telecommunication domain for various purposes such as (1) customer service [40], (2) linguistics in telecom [41], (3) autonomy to network applications in 6G (e,g. remote sensing, control planning, routing, accident detection)[42], (4) proposed new Small Language Models (SLMs) [43], and Large GenAI Models (LGMs) [44], (5) understanding telecom standards and proposed a new model which is called TeleROBERTA [45], (6) code generation in telecom [46], (7) proposed new dataset for telecom domain [47], and (8) understanding telecom language [18].

In this chapter, we explain LLM essentials:

- In Section 2.2, we give a simple definition of AI followed by a simple explanation of the term GenAI.
- The history of LLMs is explained in section 2.3.
- Next, we give an overview of LLMs in section 2.4.
- We discuss the comparison between the most used LLMs in section 2.5.
- Afterwards, we outline the tasks of LLMs in section 2.6.
- Following that, we explore the applications of LLMs in section 2.7.
- We then delve into the challenges faced by LLMs in section 2.8.
- in section 2.9 and 2.10 , we detail significant papers on Large Language Models (LLMs) in telecommunications (LLMs-Telecom) and summarize related work in a table by LLM tasks, selected models, purpose, dataset, and evaluation metric.
- Finally, we summarize this chapter in Section 2.11.

2.2 Artificial Intelligence and Generative Artificial Intelligence

Artificial Intelligence (AI) is a way of making a computer, a computer-controlled robot, or a software think and behave intelligently, in the way that intelligent humans process [48].

Researchers have been working for a long time to enable machines to read, write, and communicate like humans [37].

Generative Artificial Intelligence (GenAI) represents a close copy of the human brain's thinking to complete the required tasks. GenAI is enhancing a major transformation in the communications field, in which the exploitation of GenAI models in wireless networks or other telecom applications to handle different tasks eliminates the need to train and build AI models [49].

According to [48], GenAI is a part of AI, and it is the system that generates new content (text, images, audio, videos). Chatbots and code generators are examples of GenAI systems. Therefore, GenAI can open a new beginning in the autonomous field of wireless networks.

In the digital age, large language models have revolutionized how we understand and produce languages [39], being a crucial part of generative AI, are transforming language understanding and production in Natural Language Processing (NLP) by processing large amounts of text and accurately predicting the next word in a sentence based on previous words.

2.3 LLMs History

According to [48, 12], we summarize the history of LLMs in these points:

- 1950s and 1960s: The first language models made in the 1950s and 1960s used rules and hand-crafted features to understand language, but they couldn't handle complex language tasks.
- 1980s and 1990s: In the 1980s and 1990s, statistical language models emerged, estimating word sequences' likelihood in context with probabilities, handling more data and offering better accuracy than rule-based ones. Yet, they still had trouble fully grasping language semantics and context.
- 2010s: In the mid-2010s, neural language models marked a significant advancement, employing deep learning to comprehend language patterns from vast text data. The first of these models, the Recurrent Neural Network Language Model (RNNLM), introduced in 2010, excelled in capturing word context, generating text that sounded more natural than its predecessors.
- 2015: Google launched the groundbreaking Google Neural Machine Translation (GNMT) system, the first large-scale neural language model. Trained on extensive bilingual text data, it attained cutting-edge performance in machine translation tasks, marking a significant milestone in the evolution of LLMs.
- 2017: LLM development progressed further with the debut of the Transformer model in 2017. It excelled in learning longer-term language dependencies and enabled parallel training on multiple Graphical Processing Units (GPUs), facilitating the training of significantly larger models.
- 2018: OpenAI's release of first Generative Pre-trained Transformer (GPT-1) in 2018 signaled a major advancement in NLP, showcasing the potential of transformer-based architecture with 117 million parameters. Despite its limitations, it set the stage for subsequent,

more powerful models and sparked intense competition in LLM research, ushering in a new era of AI exploration in natural language processing.

- 2020: OpenAI introduced GPT-3, the largest language model ever created. Trained on an extensive corpus of text data, GPT-3 excelled in generating remarkably coherent and natural-sounding text, showcasing the immense potential of LLMs across various NLP applications.

2.4 General overview of LLM

- **Definition of LLM**
  - LLMs are deep learning algorithms designed to perform various language-related tasks such as recognizing, summarizing, translating, predicting, and generating content [50].
  - According to [11], the LLMs are based on early pre-trained neural language models, and LLMs contain tens to hundreds of billions of parameters.
  - According to [48], LLMs are a form of generative AI created to produce human-like language based on a given prompt. These models are trained on vast amounts of textual data from different sources including books, social media, text from poetry, songs, research articles, news articles, etc.
  - The LLM is the result of the fusion of two key fields: NLP and AI [51]. This combination has led to the development of advanced language models capable of understanding and generating human-like text [52].
  - LLMs are a type of AI algorithm that can perform a wide variety of NLP tasks [12].
  - We can use LLMs in two ways: (1) to generate desired outputs for a variety of tasks after training them, and (2) directly through basic prompting [11].

According to [11], Fig.2.1 represents the LLM families (GPT, Large Language Model Meta AI (LLaMA), and Pathways Language Model (PaLM).

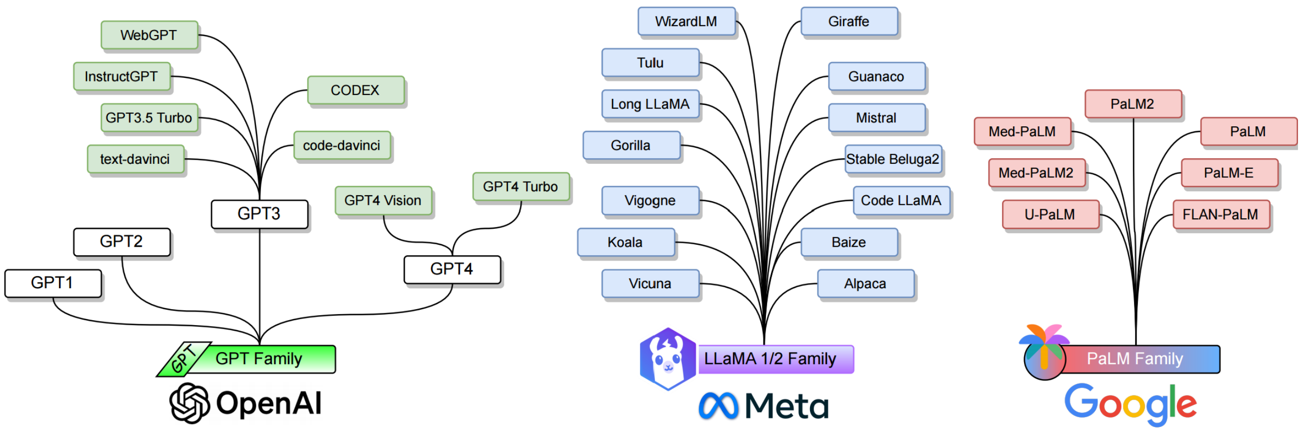

Figure 2.1: LLMs Families [11].

- **The Architecture of LLM**

  According to [11], LLM can be categorized into various classes based on the type of transformer, and each has a unique set of tasks:

  – Encoder Only: The encoder takes tokenized text (converted into numerical values) and creates meaningful representations of these tokens to group similar words together in a vector space. Models like Bidirectional Encoder Representation from Transformer (BERT) are great for tasks that involve understanding languages, such as classification and sentiment analysis.

  – Decoder Only: The decoder converts tokens back into human-readable words. After training, LLMs can predict the next words in a sequence, allowing them to perform tasks like answering questions, translating languages, and searching for meanings. Models like GPT-3 excel at generating language and content, making them ideal for tasks such as story writing and blog generation.

  – Encoder-Decoder models, such as Text-to-Text Transfer Transformer (T5), integrate language comprehension and generation, making them ideal for tasks like translation and summarization.

  The architecture of LLM is based on the attention mechanism: These are algorithms used in LLMs that help the model focus on specific parts of the input text, connecting related words [53].

  LLM architecture takes text data from different sources and sends it for preprocessing. Then, it trains by going through several steps: starting with random parameters, taking in numerical data, calculating losses, optimizing parameters, and repeating the training. After training, it can translate text, summarize text, analyze sentiment, and provide other services [12].

  According to [12] and as shown in Fig.2.2, the LLM architecture is illustrated as follows:

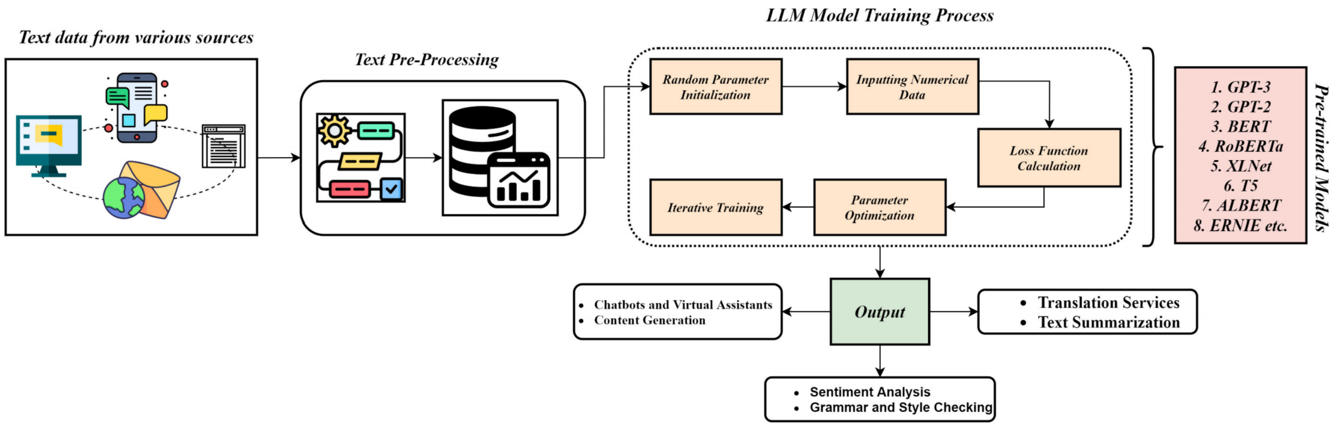

Figure 2.2: LLM Architecture [12].

- **Steps to create LLMs**

  In this section, we explain how LLMs are created briefly, the steps are represented as follows:

  – Step 1: Data cleaning: Data quality impacts the performance of LLMs [11]. The most common techniques of data cleaning are: removing noise, handling outliers, deduplication, addressing imbalances, text preprocessing, and dealing with ambiguities.

  – Step 2: Tokenizations: Tokens are produced by converting a text sequence into smaller segments [11].

  – Step 3: Positional Encoding: According to [11], this step contains: Absolute Positional Embeddings: (APE), Relative Positional Embeddings: (RPE), Rotary Position Embeddings, and Relative Positional Bias

  – Step 4: LLM Architecture: There are three main types of LLMs according to their architecture which are (1) encoder-only, (2) decoder-only, and (3) Encoder-Decoder.

  – Step 5: Model Pre-training: The model is initially trained on a large and diverse dataset to learn general language patterns, grammar, syntax, and semantics[41].

    Pre-training is the initial step in training LLMs . It helps LLMs gain basic language understanding, useful for various language-related tasks. During pre-training, the LLM is trained on a large amount of unlabeled text, typically in a self-supervised way [11].

  – Step 6: Fine-tuning and Instruction Tuning: After the pre-training phase, the model is further trained on a more specific dataset or task. This process allows the model to adapt its general knowledge to the nuances and requirements of the particular domain or task [41].

For the foundation model to be useful, it must be fine-tuned for a specific task using labeled data, a process known as supervised fine-tuning (SFT)[11]. A key reason for fine-tuning LLMs is to ensure their responses align with human expectations when given instructions through prompts[11].

  - Step 7: Alignment: AI alignment is the process of guiding AI systems to adhere to human goals, preferences, and principles. LLMs, which are pre-trained for word prediction, can often display unintended behaviors. For example, they might generate content that is toxic, harmful, misleading, or biased [11].

- **Key Factors in Model Performance**

  According to [54, 48], the performance of a model does not depend only on its size, there are three parameters to play with which are:

  1. The larger the model, the more effective it is, i.e. Larger models have more parameters, enabling them to learn more complex relationships between words and concepts.
  2. The size of the training dataset used is significant, i.e. if the dataset is too small, the model may fail to learn how to generalize to new situations.
  3. And the quality of the training dataset, i.e. if the dataset contains biased or inaccurate information, the model will learn to replicate that bias or inaccuracy.
  4. Another factor is the model's architecture, as some architectures are better suited for certain tasks than others.
  5. The task on which the model is being evaluated.

2.5 Comparison between the most used LLMs

We summarize the difference between the most used LLMs in Table2.1.

LLM Models	versions	Released By	Parameters	License	tokens
GBT-4	GBT4 Turbo GPT4 Vision	Open AI	1.6 T	Proprietary	32,768
GBT-3	GPT-3.5 GPT-3.5 Turbo	Open AI	175B	Proprietary	2048
BLOOM	560m 1B 3B 7B	BigScience	176B	open-source	250k
Noor	1B	TII	10B	open-source	/
Falcon-180B	3B	TII	180B	Apache 2.0	2048
Gemini-pro (BARD)	7B	Google	137B	Proprietary	32K
BERT	176B	Google	345M	Apache 2.0	512
Palm	Med-pal palm-E Palm-2	Google	540B	proprietary	8K
Lamda	/	Google	173B	proprietary	/
FLAN UL2	Flan-T5	Google	20B	Apache 2.0	1024
LLAMA2-70B	7B 13B	Meta AI	70B	open source	4096
MPT	7B	MosaicML	30B	Apache 2.	80K
Mistral 8x7B	Mistral 7B	Mistral AI	7.3B	Apache 2.0	32K

Table 2.1: Comparison between the most used LLMs [16]

2.6 The tasks of LLMs

As LLM has developed, they can do more things. According to [50, 55, 40, 41, 43, 46, 17], LLMs are mainly used for text-related tasks, which can be divided into the following categories:

- Content Creation: writing stories, marketing content creation.
- Software Testing.
- Semantics abilities and generative abilities.
- Summarization: rephrasing legal content and summarizing meeting notes.
- Translation: translation of languages and reading comprehension.
- Code Generation and common-sense reasoning.

- Classification: Identifying toxic content and analyzing sentiment.
- Question-answering.
- Chatbot: helping as virtual assistants.

2.7 Applications of LLMs

LLMs have various applications due to their advanced capabilities. According to [48, 12, 13, 56], the most popular LLM applications are as follows:

1. Medical domain (healthcare): There are many tools in education, radiologic decision-making, clinical genetics, patient care, etc. like XrayGPT (It can be employed for automated analysis of X-ray images and enable users/patients to ask questions about the analysis. Through the conversations, users can gain insights into their condition via an interactive chat dialogue).

2. Education: LLMs have the potential to revolutionize various aspects of learning, teaching, and educational research in the education sector. XLNet helps understand texts and documents, which can be beneficial in the academic sector. Additionally, other models can make the education system more engaging, accessible, and productive for both students and teachers.

   The influence of AI on education has been extensively debated in recent years. One notable area where AI is making a considerable impact is in student assignments and exams (CodeGPT An AI assistant to find errors in code, debug code, and more) LLMs are used in the education domain by:

   - Tutoring and Educational Resources: LLMs can be utilized to design individualized learning opportunities by providing tutoring and ideas for comprehending difficult subjects.
   - Academic Research: Within academic research, LLMs play a crucial role in conducting literature reviews, analyzing data, and aiding in the creation of research papers or reports.

3. Finance: BloombergGPT A Large Language Model for Finance [48].

4. Engineering-related applications: Software engineering (TexGPT Harnesses GPT-3's power to help you write in Overleaf, AgentGPT Autonomous AI agent in the browser), mechanical engineering, Mathematics, manufacturing [48].

5. Automating customer service: This case can be used in various industry domains, including e-commerce, banking, financial services, healthcare, etc. Automating customer service can be done by:

- Chatbots and Virtual Assistants:This means dialogue generation, that is, LLMs use advanced chatbots and virtual assistants that offer human-like interactions to help customers answer questions, provide support, and provide information.
- Answering customer questions: It is based on a database that can provide the most relevant needs for customers.
- Automate email and social media responses: This saves time and costs. It can be helpful in many ways, including sales and marketing.

6. Generate content: The primary goal of content generation is to increase workers' productivity and efficiency. The generated content includes:
    - Automated journalism: Creating content such as articles, blogs, and posts depends on the demands of the journalists.
    - Creative writing: In the realm of creativity, LLMs can assist in crafting stories, dialogues, ideas, or longer works of fiction based on the prompt and the writer's objectives.
7. Language translation: This task can be:
    - Multilingual Translation: LLMs improve accuracy and speed, offering high-quality translations in various languages for documents and websites. This capability is especially beneficial for international companies and multilingual individuals as it helps to break down language barriers.
    - Localization Services: LLMs can adapt content for specific cultures and regions beyond simple translation, ensuring that the translations are accurate and culturally appropriate.
8. Data Analysis and Business Intelligence:
    - Analyzing sentiment: LLM determines the sentiment of the public about different products, services, and topics by analyzing customer reviews, social media posts, and market trends.
    - Intelligence business: LLMs are capable of extracting complex information (papers, articles, meetings, etc.) into understandable business reports, market analyses, and executive summaries. They help financial advisors summarize earnings calls and record crucial meetings. LLMs can analyze and summarize legal or financial documents, extracting important information for clear presentation and decision-making. LLMs are valuable tools for collecting and analyzing market data, essential for developing content strategies and starting new products.
9. Telecommunications: According to [54], LLM and 6G networks offer great opportunities. LLMs can change how 6G networks work, making them smarter. Plus, 6G can link lots of LLM agents together.
    - Smart Sensing with LLMs in 6G: In 6G, smart sensing is crucial, and LLMs can help with that. By combining different types of sensing, such as radio signals,

images, and radar, LLMs can do things such as detect objects, track movement, and even measure health. This could change industries like smart cities, traffic management, security, and healthcare.

- LLMs in wireless communication: In addition, LLMs can use different types of sensory data to improve wireless communication. This includes tasks such as planning radio networks and controlling power. The goal is to ensure reliable connections and better performance everywhere.
- Linking numerous LLM agents through a 6G network opens the door to collective intelligence and automation.

10. Agriculture: LLMs can analyze vast data repositories containing soil, crop, and weather data, in addition to satellite imagery.
11. Law and Social media: LLMs have significantly impacted the social media industry, especially in content production, moderation, and sentiment analysis. They can perform tasks such as writing content, classifying text, and even generating full blogs and articles for social media.

Fig.2.3 represents the most important LLM applications.

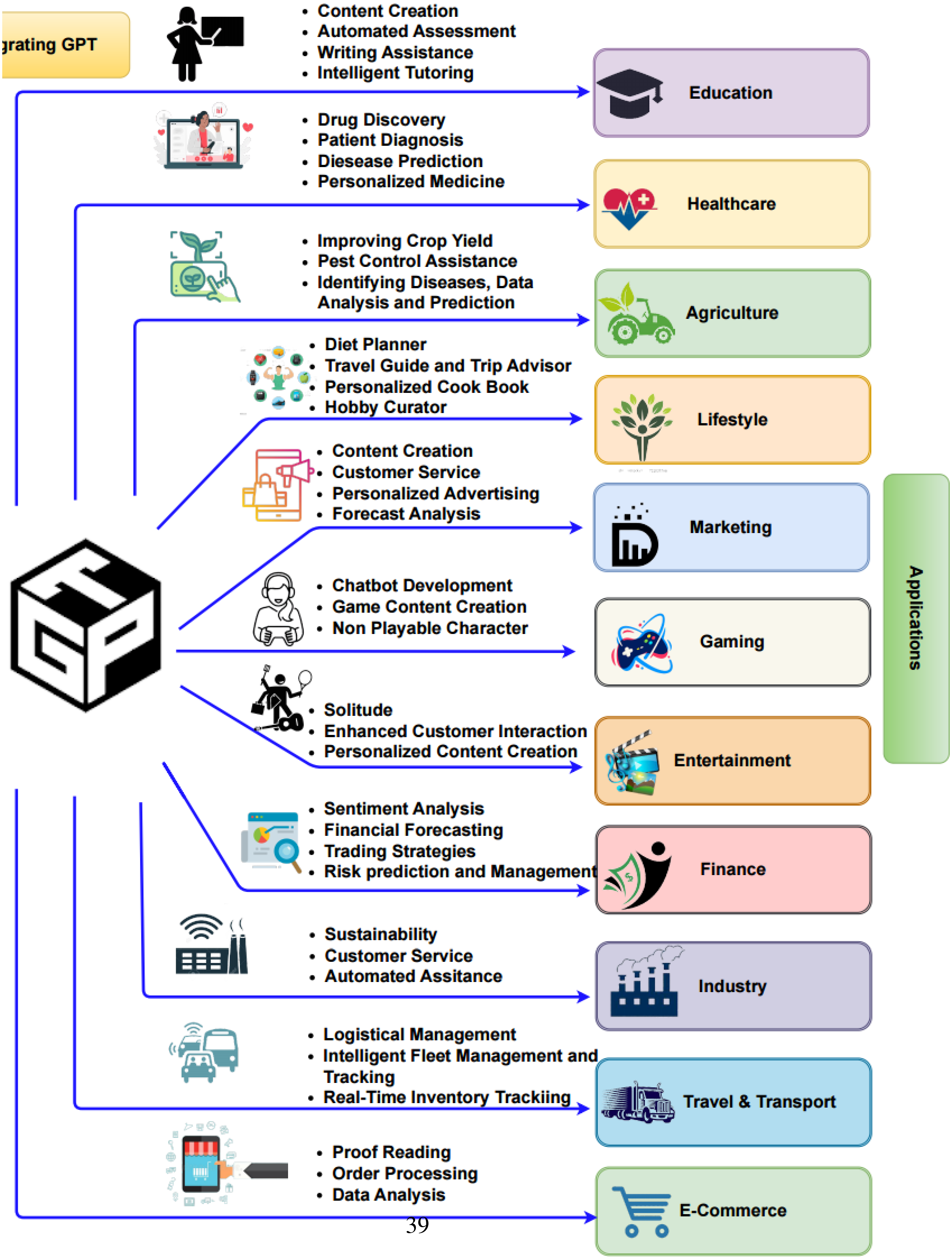

Figure 2.3: LLM Applications [13].

2.8 LLM Challenges

According to [12, 13, 56, 11], LLM challenges are:

- Unfathomable Datasets: The vastness of modern pre-training datasets makes it impractical for any individual to thoroughly read or assess all the encompassed documents.
- Tokenizer-Reliance: Tokenizers introduce various challenges, such as computational overhead, language dependency, management of new words, fixed vocabulary size, loss of information, and limited human interpretability.
- Ethical: The ethical use of large language models poses an ongoing question. Challenges persist in filtering, moderating, and ensuring accountability for AI-generated content. Issues like misinformation, hate speech, and biased content produced by LLMs underscore the need for continuous research and development.
- Multimodel: As LLMs primarily focus on text, there's a rising need for multimodal models capable of understanding and generating content that incorporates text, images, and other media types. However, integrating multiple modalities into a single model presents challenges in data gathering, training, and evaluation.
- Energy: The environmental impact of training and deploying large language models remains a pressing concern. Developing more energy-efficient training methods, model architectures, and hardware solutions is crucial to mitigate the carbon footprint associated with LLMs.
- Security: LLMs are susceptible to adversarial attacks, wherein slight modifications to inputs can result in unexpected and potentially harmful outputs. Enhancing model robustness and security against such scenarios is a critical area of research, especially for cybersecurity and content moderation applications.
- Privacy: As LLMs become more proficient, concerns regarding user privacy and data protection intensify. Finding methods for users to engage with these models without compromising their personal information poses an ongoing challenge. Research on privacy-preserving techniques and regulatory compliance is essential to address these concerns.

Fig.2.4 represents the most important LLM challenges.

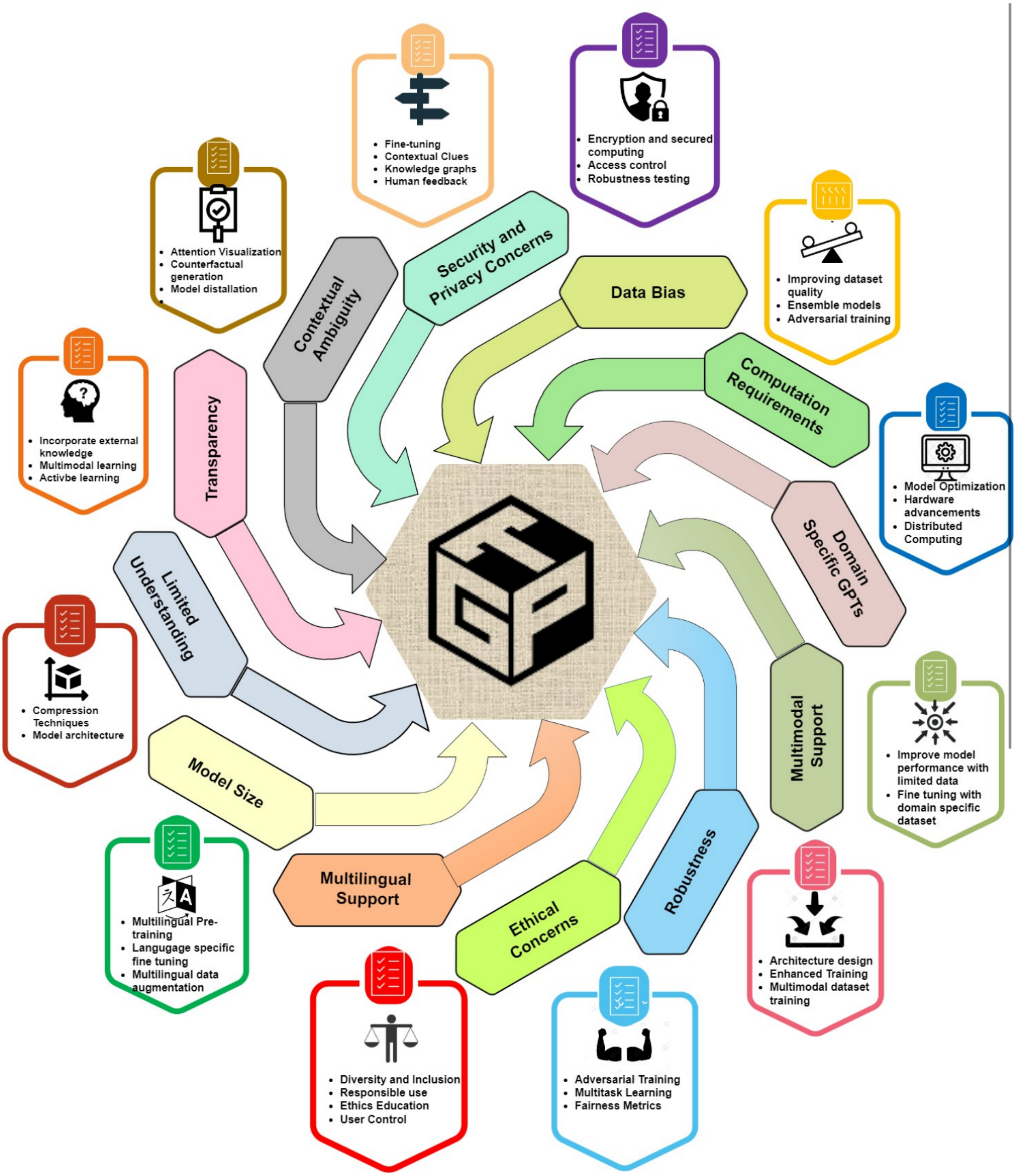

Figure 2.4: LLM challenges [13].

2.9 Related Work on Using Large Language Models in the Telecom Domain (LLMs-Telecom)

Our related work emphasizes the crucial role of LLMs in enhancing telecommunications, which include:

1: Design of a large language model to improve customer service in telecom operators [40]

- The authors highlight the current challenges of traditional customer service systems, which include long processing times, inaccurate results, the necessity to preserve proprietary knowledge bases, and the requirement to protect user privacy.
- The article discusses the design and implementation of a large-language model to enhance customer service in the telecommunications sector. Additionally, they emphasize the use of reinforcement learning to boost model performance and decrease the generation of incorrect information.

2: Linguistic Intelligence in Large Language Models for Telecommunications [41]

- The paper investigates linguistic intelligence in large language models for telecommunications, focusing on evaluating the comprehension and knowledge capabilities of LLMs in this field.
- The authors conducted a comprehensive evaluation of four prominent LLMs - Llama-2, Falcon, Mistral, and Zephyr - through text classification, summarization, and question-answering tasks.
- Zero-shot means that learning in the context of large language models (LLMs) refers to the ability of a model to perform a task without any prior specific training or examples related to that task. In other words, the model uses its general understanding of language and the world, gained from extensive training on a broad range of texts, to understand and complete a task it hasn't been specifically trained for[41].
- The results indicate that zero-shot LLMs can achieve performance levels comparable to current fine-tuned models, but still lag behind them. In particular, Zephyr emerged as the top performer in the text classification task, while BERT5G, pre-trained on the SPEC5G dataset, outperformed the other LLMs in the summarization task. In the multiple-choice question-answering task, Mistral displayed impressive performance with an accuracy of 60.93%, but all LLMs struggled with generating accurate answers in traditional question-answering.

- The results emphasize the need for further research on prompt optimization, effective training techniques, and integration of human evaluation for tasks such as traditional question-answering.

3: GenAINet: Enabling Wireless Collective Intelligence via Knowledge Transfer and Reasoning[42]

- This paper introduces the concept of GenAINet, a network based on generative artificial intelligence (GenAI) that aims to harness collective and artificial intelligence (AI) for 6th generation (6G) wireless networks.
- The paper highlights the need to enable GenAI agents to communicate knowledge to perform complex tasks via a wireless network. To achieve this, the paper proposes an architecture of the GenAINet network, focusing on semantic knowledge representation, multimodal semantic reasoning, and native semantic communication to enable GenAI agents to collaborate effectively.
- Two case studies are presented to illustrate the application of GenAINet:(1) knowledge transfer for wireless device queries and (2) wireless power control. The results show that the use of semantic communication and collaborative reasoning can improve the efficiency of GenAI agents in solving specific tasks.

4: Telecom Language Models: Must They Be Large? [43]

- This paper examines the growing interest in LLMs in the telecommunications sector, highlighting their potential to revolutionize operational efficiency. However, the deployment of these sophisticated models is often hindered by their substantial size and computational requirements, raising concerns about their viability in resource-constrained environments.
- To address this challenge, recent advances have seen the emergence of small language models (SLMs) that demonstrate comparable performance to their larger counterparts in many tasks, such as coding and common-sense reasoning. Highlights the emergence of SLMs as a potential solution, with Phi-2 from Microsoft being a notable example.
- Phi-2, a compact yet powerful model, exemplifies this new wave of efficient language models. A comprehensive evaluation of Phi-2's intrinsic understanding of the telecommunications domain is presented by comparing its performance against larger models like GPT-3.5 and GPT-4, showcasing its enhanced capabilities through a Retrieval Augmented Generations (RAG) to enhance Phi-2's capabilities, particularly in addressing telecom standards and problem-solving scenarios.
- The study utilizes the TeleQnA dataset to evaluate the language models' performance within the telecom domain, providing a nuanced assessment of their potential applications and limitations.

5: Using Large Language Models to Understand Telecom Standards[45]

- This paper discusses the use of LLMs to understand and reference Telecom Standards, specifically focusing on the Third Generation Partnership Project (3GPP) standards. The increasing volume and complexity of these standards have made it difficult for vendors and service providers to access relevant information.

- The paper evaluates the performance of LLMs as Question Answering Assistants (QA) for 3GPP documents and provides benchmarking methods, data preprocessing, and fine-tuning techniques.

- The study introduces a model called TeleRoBERTa, which performs comparably to foundation LLMs with fewer parameters. The findings suggest that LLMs can be valuable tools for referencing technical documents, facilitating troubleshooting, maintenance, network operations, and software product development.

- Experiments were conducted to evaluate the performance of different LLM models in referencing 3GPP standards.

6: Test Code Generation for Telecom Software Systems using Two-Stage Generative Model [46]

- This paper proposes a framework for automated test generation for large-scale telecom software systems, which have become increasingly complex due to the evolution of telecom towards achieving intelligent, autonomous, and open networks.

- The framework involves generating test case input data for test scenarios observed using a time-series generative model trained on historical telecom network data during field trials, along with preserving the privacy of telecom data.

- The generated time-series software performance data are then used with natural language test descriptions to generate a test script using the large-language generative model.

- The paper focuses on generating a test script with test cases for a given test description written in natural language. The method is a hybrid generative model that utilizes Generative AI for Test automation, Large Language Models for Software Testing, and Code Generation.

- The proposed framework reduces lead time in telecom software development, shows the ability to leverage historical product performance data and demonstrates its robustness and generalization through comprehensive experiments on public datasets and telecom datasets obtained from operational telecom networks.

7: TeleQnA: A Benchmark Dataset to Assess Large Language Model Telecommunications Knowledge [47]

- In this document, the term benchmark refers to a set of standards or criteria used to assess the performance or proficiency of LLMs in the field of telecommunications.
- The paper introduces TeleQnA, the first benchmark dataset designed to evaluate LLMs in the field of telecommunications.
- Comprising 10,000 questions and answers sourced from standards and research articles, TeleQnA evaluates LLMs' capabilities, such as GPT-3.5 and GPT-4, in addressing telecom-related inquiries.
- The results show that while LLMs excel in general telecom knowledge, they struggle with complex standards-related questions.

8: Understanding Telecom Language Through Large Language Models [18]

- The paper presents a study on the use of LLMs to understand the language of telecommunications, focusing particularly on the technical documents of the third-generation partnership project (3GPP).
- The authors adapted several LLM models, such as BERT, RoBERTa, and GPT-2, to the language of telecommunications by training them in relevant technical documents from the 3GPP to predict the categories of Tdoc documents for the years 2020-2023.
- The results showed that the fine-tuning of the BERT and RoBERTa models achieved an accuracy of 84.6% in identifying the 3GPP working groups, while the GPT-2 model reached 83%. The DistilBERT model, with around 50% fewer parameters, achieved similar performance, demonstrating that fine-tuning pre-trained LLMs can effectively identify telecommunications language categories.
- The document also emphasizes the importance of creating large annotated datasets from 3GPP technical specifications and highlights the impact of the length of technical text segments on classification accuracy. Experiments showed that increasing the length of text contributes to improved accuracy, but this effect diminishes as the length increases, underscoring the importance of finding a balance between performance and computational complexity for LLMs.
- Furthermore, the paper highlights the importance of developing LLMs capable of understanding telecommunications language, as a cornerstone for enabling autonomous networks driven by intelligent generative agents. The results of the experiments showed that the fine-tuned models accurately classified technical texts related to telecommunications, thus highlighting the potential of these models for future use in wireless networks.

2.10 Summary of Related Work

In this section, we give a brief overview of our related work in the table shown below (Table 2.2):

Paper	LLMs tasks	Selected LLMs	Purpose	Dataset	Evaluation metrics
[40] 2024	Question Answering (Q&A)	Chat GLM2	Customer Services	Q&A	MSE (Mean Squared Error loss) KRA (Knowledje Recommondation Accuracy)
[41] 2024	Summarization Text classification Q&A	LLama 2 Falcon Mistral Zephyr	Linguistic	SPEC5G TeleQnA	Rouge 1,Rouge 2 Rouge 3,Accuracy Preecision F1,Recall Accuracy R1, R2, LL
[42] 2024	Q&A	GPT 3,5 turbo	Autonomy in 6G network	TeleQnA	Accuracy Reduce of bits exchange (%)
[43] 2024	Coding Common sense reasoning	Phi-2 (Microsoft) GPT	Proposed Small Language Model (SLM)	TeleQnA	Accuracy Blue Energy consumption
[45] 2024	Q&A	TeleROBERTA GPT 3,5 turbo GPT 4 LLama 2 Falcon	Understanding telecom Standars and proposed a new model TeleROBERTA	TeleQuAD ([57])	Accuracy
[46] 2024	Text generation	LLama Mistral	Code generation	MBPP Humameral.X	CS(Content Score) TCER (Test Case Effectiveness Rate) Meteor
[47] 2023	Q&A	GPT 3,5 GPT 4	Proposed a new dataset	TeleQnA	Accuracy
[18] 2023	Text classification Understanding	Bert,Roberta GPT-2 DistilBert	Understanding telecom language	3GPP technical documents (Tdoc)	Accuracy

Table 2.2: Summary of Related Work.

2.11 Conclusion

Large language models (LLMs) are powerful tools for quickly and accurately handling natural language data with minimal human help. Since LLMs can quickly process large amounts of text data with precision, they are becoming essential for various uses in many industries[38].

The size of large language models (LLMs) has grown exponentially in recent years, roughly tenfold each year. As LLMs continue to expand, they will significantly shape how we create, communicate, and utilize technology in the future.

In this chapter, we discuss the latest advancements in using LLMs within the telecommunications sector (LLM-Telecom) and present a summarized view in a table to enhance comprehension of our research context. Moving forward, the next chapter will delve into our specific contributions, focusing on troubleshooting base station anomalies using LLMs, followed by detailed experimental procedures and discussions.

With LLMs at our disposal, there are countless opportunities for innovation and growth. We can accomplish a lot and explore new possibilities for development and creativity [39].

Chapter 3

Solution and Results

Contents

3.1 Introduction

In the previous chapters, we covered the background of mobile networks, large language models (LLMs), and our related work. This chapter focuses on a critical aspect of both industry and academic research.

In this chapter, our focus includes:

- Section 3.2: Detailed description of our problem statement.

- Section 3.3: Selection and discussion of key motivations driving our proposed contribution.
- Section 3.4: Brief explanation of our chosen programming language for this work.
- Section 3.5: Step-by-step explanation of our dataset cleaning and pre-processing procedures.
- Section 3.6: Implementation details of applying LLMs to generate solutions for our dataset (specifically using GPT-2, with similar procedures applied to Llama and Mistral models).
- Section 3.7: Presentation of our evaluation methodology.
- Section 3.8: Analysis and discussion of results.
- Section 3.9: Conclusion summarizing the findings of this chapter.

3.2 Problem Description

In our days, telecom has become an essential industry in the world, as the number of users is increased, it becomes important to advance this industry to ensure seamless and continuous connectivity.

However, many disturbances in the equipment and BSs can disrupt the network. The main problem we face is the traditional way of handling these anomalies, which may not be sufficient to troubleshoot the complexity and scale of modern telecom networks.

During our internship with Mobilis Maintenance in Setif, Algeria, we observed that there are 370 BSs (2G, 3G, 4G) in the area. The Mobilis maintenance office provided us with access to a dataset through their 4T platform. This national platform, established in 2010 for the Mobilis agency, allows each technician to log in using their unique username and password. The 4T platform can be accessed via PCs or smartphones and serves as a trouble report system for BSs, containing sensitive information pertinent to maintenance issues, which are kept confidential within the agency. We analyzed the problem in detail, focusing on four key elements: a technician in the field, a maintenance head in a specific region, the head of the national Mobilis agency, and the BS. The process is as follows:

1. Each week, the maintenance head in a specific region submits a list of technicians scheduled to work that week.
2. The head of the national Mobilis agency updates the 4T platform accordingly.
3. When an alarm is triggered, the BS sends the alarm to the 4T platform.
4. The national head updates the platform with the alarm details and informs the selected technician and regional maintenance head.

5. The technician addresses the issue at the BS.

6. Once resolved, the technician reports back to the maintenance head, who then updates the 4T platform (the traditional method). The national head must verify the resolution.

7. If unresolved, the technician may seek assistance from colleagues or reference documents, especially if inexperienced, which can be time-consuming and resource-intensive.

In summary, the 4T platform does not provide direct solutions for BS alarms; technicians must resolve issues based on the information available on the platform. Our approach involves using LLMs to generate solutions, offering several benefits:

- Creating more resilient and efficient communication networks.
- Enhancing safety and reducing disaster impacts.
- Minimizing downtime and financial losses.
- Assisting technicians in resolving BS alarms.
- Protecting mobile network resources.
- Reducing resolution time

3.3 Motivations

We can select our motivations to propose a new contribution in these points:

1. The "paper Large-Language Models for Telecom: Forthcoming Impact on Industry in [17]" examines the potential impact of LLMs on the telecommunications industry. The results of this paper show that LLMs can help address network anomalies, understand 3GPP standard specifications, and model networks, but also have limitations such as hallucinations, computational complexity, and inconsistent results.

2. The base station is the cornerstone of the telecommunication infrastructure in mobile networks, and any issues related to it can result in network failure.

3. In Algeria, telecom companies such as Telecom Algeria, Djezzy, Ooredoo, and Mobilis handle base station issues using traditional methods. For instance, when an alarm signals a problem, their staff rely on their experience to solve it. However, this approach has its downsides. It could result in the base station being inactive for some time, which costs both time and money. Additionally, it may cause missed calls or slow internet for users, impacting their satisfaction with the service.

4. Based on our related work, there have not been any papers that specifically address solving base station problems using LLMs.

5. Integrating LLMs for base station troubleshooting is a promising concept worldwide. It has the potential to optimize both software and hardware resources for telecommunication operators, leading to savings in time and money. In addition, this approach would streamline the workflow for employees, making their tasks more efficient and effective.

3.4 Programming language

We chose Python as our programming language in our implementation because Python is a powerful and easy-to-use programming language that can be used for many tasks such as web development, data analysis, and machine learning, It is known to be simple and easy to read, as it is illustrated in Fig.3.1 [58].

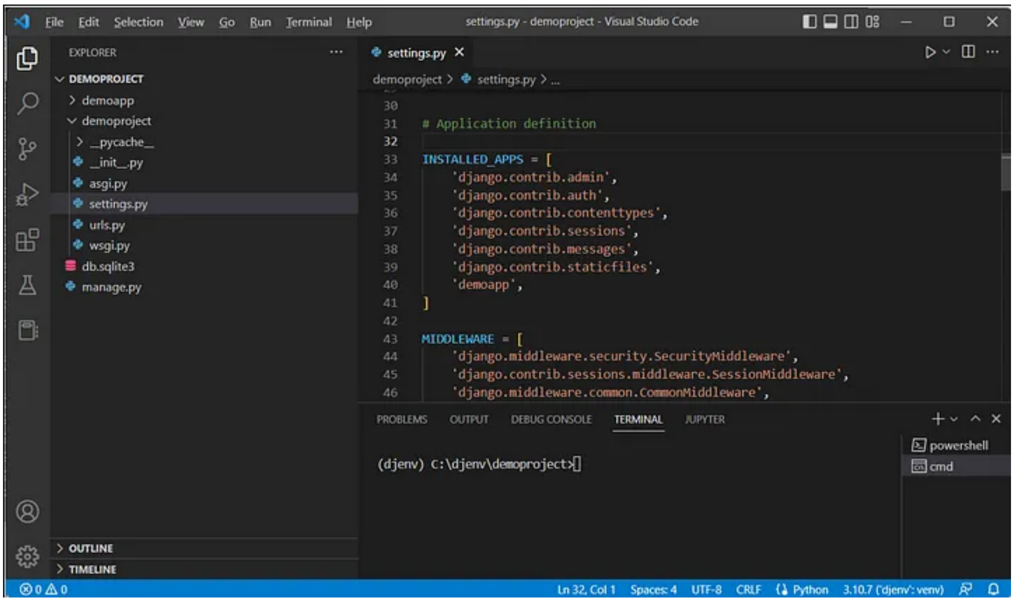

Figure 3.1: Interface python 3.10.7 [14].

Google colab https://colab.research.google.com/ is a free online tool that allows us to analyze data and do machine learning tasks [59]. Google Colab is used also to write and run Python code in your web browser as shown in Fig. 3.2.

In our case, One of the key benefits of using Colab is that we can easily use large models, without having to download the models locally on our computer, so this saves time and improves the computer's performance.

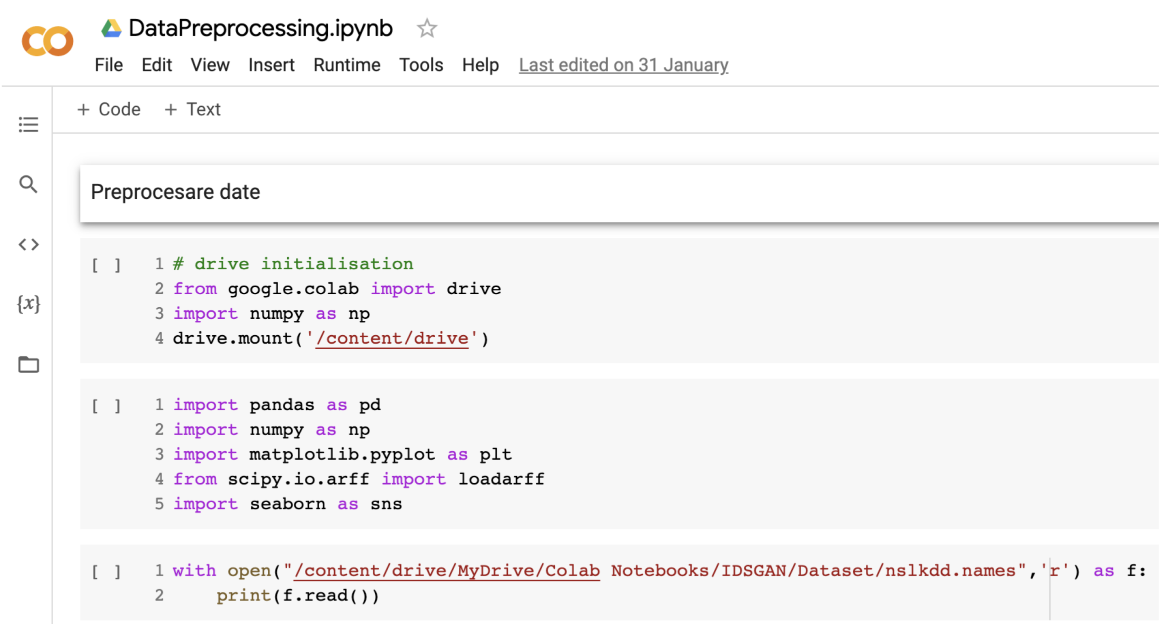

Figure 3.2: google colab interface[15].

3.5 Dataset cleaning and pre processing

This subsection outlines the steps in dataset cleaning that we will perform on the collected data set (4T) to create our LLM-TBSA dataset ultimately.

- **Definition of a dataset:**
  - According to [60, 61], a dataset is a collection of data that is used to train or test a machine learning model; a machine learning model is an algorithm that learns patterns from the data to make predictions or decisions based on new input.
  - Data can be categorized into two main types, structured and unstructured. Structured data are organized in a two-dimensional table format, while unstructured data do not have a predefined structure.
  - Within these categories, datasets can be further classified according to their content. Text datasets consist of text data, image datasets contain images, and voice datasets include voice data.
  - The main purpose of a dataset in machine learning is to provide examples for the model to learn from and make predictions or decisions about new data [62].
- **Steps to download a dataset**

  According to [63, 64], Here are some of the most popular websites for finding datasets:

  1. Kaggle https://www.kaggle.com/.

2. Google Dataset Search https://datasetsearch.research.google.com/.
3. Hugging Face https://huggingface.co/docs/datasets/index
4. FiveThirtyEight https://data.fivethirtyeight.com/.
5. Data.gov https://data.gov/
6. Datahub.io https://github.com/
7. AWS Public Datasets https://registry.opendata.aws
8. data.world https://data.worldbank.org

- **4T dataset**

  We can say that the Mobilis agency gave us a simple dataset, we call it a 4T dataset (the original one). The 4T dataset contains 4264 rows and 16 columns.

The columns are as follows:

- TT ID: Trouble Ticket IDentifier, an identifier for each anomaly,e.g.1085103.
- Titre: Title of each anomaly,e.g.18656 TRX DEGRADED.
- Statut: Status of the anomaly,e.g.En cours (Attribué).
- Create Date: Time of the alarm,e.g.06/03/2024 06:20:00.
- Catégorie: Category of the anomaly, such as power, transmission, or radio, followed by the specific anomaly,e.g.FMD Power > External Power.
- Attribué à - Groupe de techniciens: Group of technicians responsible for the specific site where the BTS is located,e.g.FMO > UOP-Tindouf.
- Technicien: Name of the technician, e.g.Kebir Mouaadh.
- Types d'élément associé: Principal state the BTS belongs to,e.g.Bechar.
- éléments associés: Associated elements, e.g. N_1983 / LTE_TINDOUF -.
- Priorité: Risk level of the anomaly (high, medium, low),e.g.Haute.
- Solution: Solution to the anomaly.
- Affected Service: Service affected by the anomaly either "BTS Service Degradation" (reduced performance) or "BTS Down" (completely out of service),e.g.BTS Totaly Down.
- Description: Description of the anomaly, e.g.37999 NodeB is out of service.
- Network elements type: Type of network element containing the problem,e.g.Zte_NodeB.
- DMR - Type du Site: Type of site (normal or HUB),e.g.Normal.
- Dernière modification: Date and time of last modification,e.g.06/03/2024 06:27:00.

3.5.1 Dataset Cleaning Steps

1. We remove the page header shown in Fig.3.5.

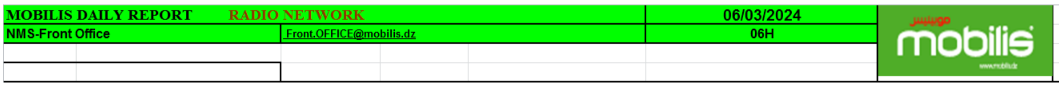

Figure 3.3: The page header of 4T dataset.

2. We convert the Excel document to a CSV file (see Fig.3.4) because it is simple and Compatible for use in Python, and it is easy to use with data analysis libraries like **pandas**.

3. We notice that the first two rows represent the columns' names, so we can make it one row it will be better.

4. We read our dataset (we have to read our dataset to complete the dataset cleaning). We upload our dataset to Google Colab as follows :

```
from google.colab import files
uploaded = files.upload()
```

5. We import pandas library to handle various data manipulation and make the work easier in Python :

```
    import pandas as pd
```

6. The command converts the file telecom Dataset.csv, which is encoded in latin1, to a new file called telecom_utf8.csv that is encoded in utf8.

```
   !iconv -f latin1 -t utf8 Telecom\ Dataset.csv >
   Tel_Dataset.csv
```

7. We read the CSV file named Tel_Dataset.csv

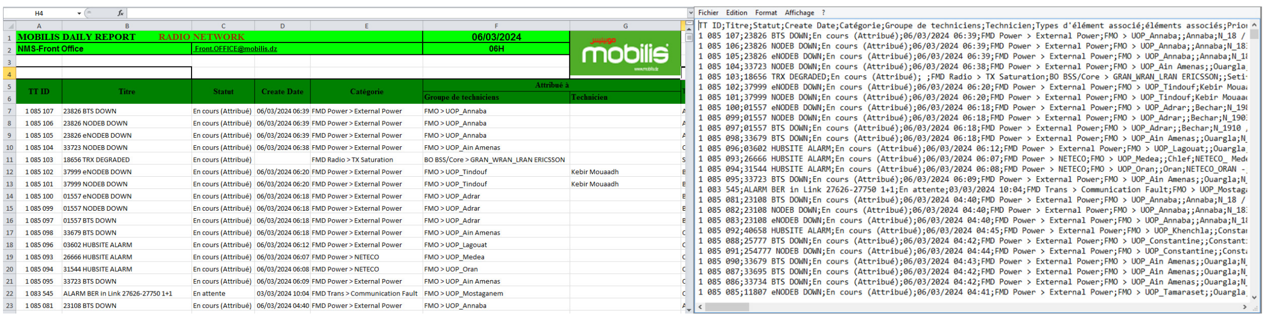

Figure 3.4: Comparison between the Excel and CSV versions.

```python
df = pd.read_csv('Tel_Dataset.csv', delimiter=';', encoding='utf-8')
```

8. We display the first 5 rows of our dataset to understand the structure and content.

```python
print(df.head())
```

9. We split the title column and display the columns to check it.

```python
df[['BS ID', 'Alarm message']] = df['Titre'].str.split(' ', n=1, expand=True)
df.columns
```

10. We omit unnecessary columns : There are some columns we do not need, we omit them and display columns to check it, as follows :

```python
columns_to_omit = ['TT ID','Create Date','Titre', 'Derni re modification','Groupe de techniciens','Technicien']
df = df.drop(columns=columns_to_omit, axis=1)

df.collumns
```

11. We omit symbols, we load two tools for working with text. One tool helps find patterns in text, the other helps handle text from different languages.

```python
import re
import unicodedata
```

12. We remove non-printable characters:

```python
def clean_text(text):
    return re.sub(r'[\x00-\x1F\x7F-\x9F]', '', text) if isinstance(text, str) else text
```

13. We have to normalize Unicode characters:

```python
def normalize_text(text):
    return unicodedata.normalize('NFKC', text) if isinstance(text, str) else text
```

14. We remove excessive punctuation :

```python
def basic_cleaning(text):
  return re.sub(r'[^\w\s]', '', text) if isinstance(text, str) else text
```

15. We convert text to lowercase:

```
def to_lowercase(text):
  return text.lower() if isinstance(text, str) else text
```

16. We remove specified symbols:

```
def remove_specified_symbols(text):
    symbols_to_remove = {'_', '  ', '  ', '  ', '  ', '
     '}
    for symbol in symbols_to_remove:
        text = text.replace(symbol, '')
    return text
```

17. We apply all cleaning functions:

```
def clean_all_text(text):
    text = clean_text(text)
    text = normalize_text(text)
    text = basic_cleaning(text)
    text = to_lowercase(text)
    text = remove_specified_symbols(text)
    return text

```

18. We apply the cleaning functions to all columns:

```
for column in df.columns:
    df[column] = df[column].astype(str)
for column in df.columns:
    df[column] = df[column].apply(clean_all_text)

```

19. Finally, we display all dataset to check if all codes work well:

```
 df

```

20. We translate all the dataset content: we notice that our dataset columns are a mixture of two languages, French and English. So, we have to translate it into English. First, we have to install the **transformers** library.

```
 !pip install transformers
```

21. We import the imports the **pipeline** function from the **transformers** library:

```
        from transformers import pipeline
```

22. We use this code for translating to English:

```
        def translate_to_english(text):
    translated_text = GoogleTranslator(source='auto', target='en').translate(text)
    return translated_text
```

23. We precise the columns that need translation :

```
     columns_to_translate = ['Statut', 'Priorit ', 'DMR - Type du Site']
```

24. The code initializes a translation pipeline using the "Helsinki-NLP/opus-mt-fr-en" model from the Hugging Face Model Hub. This specific model translates text from French to English, then we display the dataset to check if the content is translated.

```
     source_lang = 'fr'
target_lang = 'en'

translator = pipeline('translation', model='Helsinki-NLP/opus-mt-fr-en', src_lang='fr', tgt_lang='en')

for column in columns_to_translate:
  df[column] = df[column].fillna("")
  df[column] = df[column].apply(lambda text: translator(text)[0]['translation_text'])

  df
```

25. We translate column names:

```
df.columns = [translator(name)[0]['translation_text'] for name in df.columns]
```

26. We modify the name of column 'DMR - Site type', and we display columns names to check it :

```
old_column_name = 'DMR - Site type'
new_column_name = 'Site type'
df = df.rename(columns={old_column_name: new_column_name})
df.columns
```

27. We save the cleaned dataset :

```
df.to_csv('Dataset.csv', index=False)
```

3.6 Solutions Generating

Based on our related work, we selected three large language models (LLMs) for the text generation task: GPT-2, Llama, and Mistral.

In this section, we explain how we generate solutions for BS anomalies using GPT-2. The same approach is applied for generating solutions using the Llama and Mistral models.

1. We read the dataset that we saved before

```
path="/content/drive/MyDrive/Dataset/Dataset.csv"
df = pd.read_csv(path)

```

2. We omit some columns, we do not need them in solutions generating.

```
columns_to_omit = ['Status','Associated elements', '
    Description','BS ID',]
df = df.drop(columns=columns_to_omit, axis=1)

```

3. We display columns to confirm.

```
        df.columns

```

4. We upload the necessary library for execution.

```
        !pip install transformers torch nltk rouge-score

```

5. The code generates two solutions using the GPT-2 model:

```
    from transformers import GPT2LMHeadModel,
    GPT2Tokenizer
import pandas as pd
import torch
import nltk

# Define the prompt template
prompt_template = """
I work in the telecom industry. Please provide a
    solution for the base station anomaly described below
    :
- Site type: '{site_type}'
- Related Element Types: '{related_element_type}'
```

```
- Category: '{category}'
- Priority: '{priority}'
- Assigned Service: '{assigned_service}'
- Network elements type: '{network_elements_type}'
- Alarm message: '{alarm_message}'
Answer:
"""

# Initialize GPT-2 model and tokenizer
model_name = 'gpt2'
tokenizer = GPT2Tokenizer.from_pretrained(model_name)
model = GPT2LMHeadModel.from_pretrained(model_name).to('cuda')  # Move the model to GPU

# Function to generate solution for a single row
def generate_solution(row, temperature=0.7):
    try:
        # Fill in the prompt template with row values
        prompt = prompt_template.format(
            site_type=row['Site type'],
            related_element_type=row['Related Element Types'],
            category=row['Category'],
            priority=row['Priority'],
            assigned_service=row['Assigned Service'],
            network_elements_type=row['Network elements type'],
            alarm_message=row['Alarm message']
        )

        # Tokenize the prompt
        input_ids = tokenizer.encode(prompt, return_tensors='pt').to('cuda')  # Move input_ids to GPU

        # Generate text based on the prompt using GPT-2
        output = model.generate(
            input_ids,
            max_length=200,
            num_return_sequences=1,
            no_repeat_ngram_size=2,
            temperature=temperature,
            top_k=50,
```

```
            pad_token_id=tokenizer.eos_token_id,
            do_sample=True
        )

        # Decode the generated text
        generated_text = tokenizer.decode(output[0], skip_special_tokens=True)

        # Extract the answer part after "Answer:" (assuming the generated text follows the prompt)
        answer_start = generated_text.find("Answer:") + len("Answer:")
        answer = generated_text[answer_start:].strip()

        # Fallback if the model does not generate a proper response
        if not answer or answer.startswith("The base station anomaly is:"):
            answer = "The base station anomaly is:"

        return answer
    except Exception as e:
        print(f"Error generating solution: {e}")
        return "Error generating solution."

```

6. We display the dataset to confirm that all rows are generated.

```
    df.columns

```

7. We save our data set with GPT2 solutions.

```
    df.to_csv('Dataset solution by GPT2.csv', index=False)

```

These are two examples of how GPT 2 model generates two solutions for row number 1000:

1.
```
    specific_value = df.loc[1000, 'Solution_1']
print(f"Content at row for solution 1 {1000}, column '{8}': {specific_value}")

```

Solution 1: important right information cannot blindly trust person says oh saw signal need able communicate network get data phone matter data bad guy systems bad guy said paul sure security measures however comes frequency signal.

2.
```
    specific_value = df.loc[1000, 'Solution_2']
print(f"Content at row for solution 2 {1000}, column
   '{9}': {specific_value}")

```

Solution 2: scenario station least one uplink point two alarms uplinks reach point time next 5 seconds may monitored alarm case simply set dedicated antenna default antenna uplinking station add additional points needed four antenna types connected system frs sessas pts.

3.7 Evaluation

In this section, we evaluate the performance of the selected LLMs. We select a similarity metric for calculating the similarity between the two solutions generation for the same model.

To calculate this metric, we follow these steps:

1. To install the sentence-transformers and contractions libraries in Python, you can use:

```
!pip install sentence-transformers contractions
```

Sentence-transformers: This library provides an easy-to-use interface for computing sentence and text embeddings using various pre-trained models, including BERT, RoBERTa, and others.

Contractions: This library is used to expand contractions in English text. For example, it will expand "don't" to "do not" and "I'm" to "I am". This can be useful for text preprocessing in NLP tasks.

2. We clean our solutions (solution 1 and solution 2):

```
    # lowercase
df['Solution_1']=df['Solution_1'].apply(lambda x: x.
   lower())

# remove contractions (she'll => she will)
df['Solution_1'] = df['Solution_1'].apply(
   remove_contractions)

# remove punctuation (?!,)
df['Solution_1'] = df['Solution_1'].apply(
   remove_punctuation)
```

```

# remove stop words
df['Solution_1'] = df['Solution_1'].apply(lambda x: ' '.join([word for word in x.split() if word not in (stop)]))

# lowercase
df['Solution_2']=df['Solution_2'].apply(lambda x: x.lower())

# remove contractions (she'll => she will)
df['Solution_2'] = df['Solution_2'].apply(remove_contractions)

# remove punctuation (?!,)
df['Solution_2'] = df['Solution_2'].apply(remove_punctuation)

# remove stop words
df['Solution_2'] = df['Solution_2'].apply(lambda x: ' '.join([word for word in x.split() if word not in (stop)]))
```

3. We load the BERT model. The model is now ready to be used to transform sentences into their vector (embedding) representations, which can be used for various natural language processing tasks such as semantic similarity.

```
model = SentenceTransformer("bert-base-uncased")
```

4. This code calculates the cosine similarity between sentence embeddings using a pre-trained model from the **sentence-transformers** library and the cosine similarity function from **sklearn**:

```
from sklearn.metrics.pairwise import cosine_similarity

def calculate_sentence_embeddings(text):
    # Assuming `model.encode()` returns the sentence embeddings as a numpy array
    embeddings = model.encode([text])  # Ensure input is a list for batch processing
    return embeddings

def calculate_similarity(sentence_embedding1, sentence_embedding2):
```

```
    # Reshape embeddings if needed (ensure they are 2D arrays)
    sentence_embedding1 = sentence_embedding1.reshape(1, -1)  # Reshape to 2D array
    sentence_embedding2 = sentence_embedding2.reshape(1, -1)  # Reshape to 2D array

    score = cosine_similarity(sentence_embedding1, sentence_embedding2)
    return score[0, 0]  # Extract similarity score from 2D array result
```

5. We calculate sentence embeddings for two solutions from our dataset, and then we compute the cosine similarity between these embeddings:

```
embeddings1 = calculate_sentence_embeddings(df.iloc[0]['Solution_1'])
print(embeddings1)
embeddings2 = calculate_sentence_embeddings(df.iloc[0]['Solution_2'])
print(embeddings2)
similarity = calculate_similarity(embeddings1, embeddings2)
print(similarity)
```

3.8 Discussion

In previous sections, we have outlined the process of generating solutions for BS anomalies using the GPT-2 model in detail.

Similar steps were applied to generate solutions using the other selected LLMs, namely Llama and Mistral.

Our results are depicted in Fig.3.5. As evident, the Mistral model outperforms both Llama and GPT-2 models in generating solutions for BS anomalies, demonstrating superior performance in terms of cosine similarity.

3.9 conclusion

In this chapter, we provide a detailed overview of our contributions. We outline the problem statement, discuss the motivation behind our proposed solution, specify our programming language choice, and meticulously detail the steps involved in cleaning the dataset. Our primary

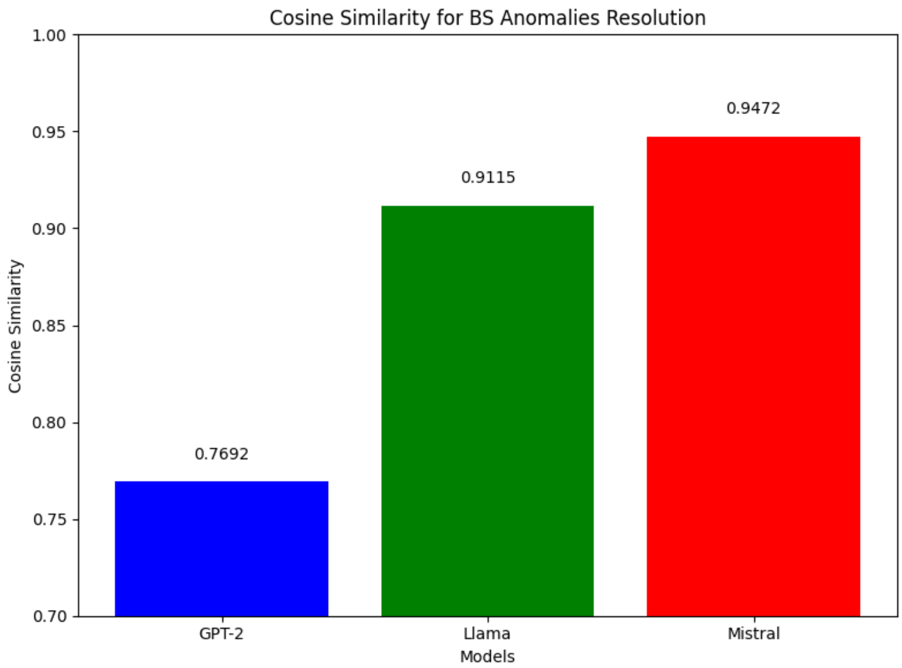

Figure 3.5: Cosine Similarity for BS Anomalies Resolution.

contribution, covered in the sixth section, focuses on leveraging the GPT-2 model to generate solutions for BS anomalies, accompanied by Python code demonstrations. We adopt cosine similarity as our evaluation metric to assess the performance of the selected LLMs, and replicate these procedures for the Llama and Mistral models. Finally, we present a comprehensive discussion of our findings.

In summary, our investigation identifies the Mistral model as the most effective LLM for generating solutions to BS anomalies.

Conclusion

1. Context

Large Language Models (LLMs) represent advanced Artificial Intelligence (AI) systems designed to process and generate human-like text using extensive training data. These models, such as GPT-3, are distinguished by their complex neural architectures, consisting of millions or even billions of parameters [11].

Our exploration of integrating LLMs into the telecommunications sector, particularly in network troubleshooting, marks a significant advancement toward a more intelligent and responsive network infrastructure. This dissertation focused on addressing base station anomalies using LLMs instead of conventional methods.

We applied LLMs specifically GPT-2, Llama, and Mistral, widely recognized in the telecommunications domain to resolve these anomalies. Our implementation and evaluation, based on cosine similarity metrics, revealed Mistral as the superior model, consistently providing more effective solutions compared to GPT-2 and Llama.

This study underscores how LLMs can transform telecom operations by automating complex tasks, enhancing accuracy in anomaly detection, and optimizing network management.

2. Futur work

- Multilingual Expansion: Currently, our solution operates exclusively in English. Plans involve expanding its applicability by incorporating additional languages such as French and Arabic.
- Broader Dataset Integration: Initially focused on Mobilis as a case study, our future endeavors aim to enrich our dataset with data from other telecom giants like Djezzy, Ooredoo, and Telecom Algeria. This expansion will generalize our findings and broaden the applicability of our approach across diverse telecommunications providers.
- Integration of Retrieval-Augmented Generations (RAGs): Incorporating RAGs into our framework aims to enhance LLMs' performance in dynamic telecom environments, leveraging contextual information for improved solution generation [42, 43, 11].

Bibliography

[1] Lillian Goleniewski. *Telecommunications Essentials: the Complete Global Source*. Jan 2007.

[2] BI Bakare and EE Bassey. A comparative study of the evolution of wireless communication technologies from the first generation (1g) to the fourth generation (4g). *Int J Elect Commun Comp Eng*, 12(3):2249–071X, 2021.

[3] Rajib Hasan. *What Is GSM Architecture in Mobile Computing System?*, Aug 2023. https://www.linkedin.com/pulse/what-gsm-architecture-mobile-computing-system-engr-rajib-hasan/.

[4] Electronics Notes. *3G UMTS Network Architecture» Electronics Notes*. https://www.electronics-notes.com/articles/connectivity/3g-umts/network-architecture.php#google_vignette.

[5] Tinatin Mshvidobadze. Evolution mobile wireless communication and lte networks. In *2012 6th international conference on Application of information and communication technologies (AICT)*, pages 1–7. IEEE, 2012.

[6] Philip Irving and Pascal A Ochang. Evolutionary analysis of gsm, umts and lte mobile network architectures. *World Scientific News*, 54:27–39, 2016.

[7] Larry L Peterson, Oguz Sunay, and Bruce S Davie. *Private 5G*. Systems Approach, May 2023.

[8] Shimaa A Abdel Hakeem, Hanan H Hussein, and HyungWon Kim. Vision and research directions of 6g technologies and applications. *Journal of King Saud University-Computer and Information Sciences*, 34(6):2419–2442, 2022.

[9] IPLOOK. *What is base station?*, May 2024. https://www.iplook.com/info/what-is-base-station-i00169i1.html.

[10] Tom. *What Is CPRI LC Outdoor Cable Assemblies FTTA?*, Jun 2024. https://mefiberoptic.com/what-is-cpri-lc-outdoor-cable-assemblies-ftta/.

[11] Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, et al. A survey of large language models. *arXiv preprint arXiv:2303.18223*, 2023.

[12] Mohaimenul Azam Khan Raiaan, Md Saddam Hossain Mukta, Kaniz Fatema, Nur Mohammad Fahad, Sadman Sakib, Most Marufatul Jannat Mim, Jubaer Ahmad, Mohammed Eunus Ali, and Sami Azam. A review on large language models: Architectures, applications, taxonomies, open issues and challenges. *IEEE Access*, 2024.

[13] Gokul Yenduri, M Ramalingam, G Chemmalar Selvi, Y Supriya, Gautam Srivastava, Praveen Kumar Reddy Maddikunta, G Deepti Raj, Rutvij H Jhaveri, B Prabadevi, Weizheng Wang, et al. Gpt (generative pre-trained transformer)–a comprehensive review on enabling technologies, potential applications, emerging challenges, and future directions. *IEEE Access*, 2024.

[14] Suhas Tumati. Project structure while working with django - suhas tumati - medium. Jun 2024. https://medium.com/@suhast_40578/project-structure-while-working-with-django-a878fd6770c4.

[15] Andrei-Grigore Mari, Daniel Zinca, and Virgil Dobrota. Development of a machine-learning intrusion detection system and testing of its performance using a generative adversarial network. *Sensors*, 23(3):1315, 2023.

[16] Deepchecks Community Blog. *LLM Models Comparison: GPT-4, Bard, LLaMA, Flan-UL2, BLOOM | Deepchecks*, Mar 2024. https://deepchecks.com/llm-models-comparison/.

[17] Ali Maatouk, Nicola Piovesan, Fadhel Ayed, Antonio De Domenico, and Merouane Debbah. Large language models for telecom: Forthcoming impact on the industry. *arXiv preprint arXiv:2308.06013*, 2023.

[18] Lina Bariah, Hang Zou, Qiyang Zhao, Belkacem Mouhouche, Faouzi Bader, and Merouane Debbah. Understanding telecom language through large language models. In *GLOBECOM 2023-2023 IEEE Global Communications Conference*, pages 6542–6547. IEEE, 2023.

[19] Lina Bariah, Qiyang Zhao, Hang Zou, Yu Tian, Faouzi Bader, and Merouane Debbah. Large language models for telecom: The next big thing? *arXiv preprint arXiv:2306.10249*, 2023.

[20] Sumit Soman et al. Observations on llms for telecom domain: capabilities and limitations. *arXiv preprint arXiv:2305.13102*, 2023.

[21] Khaled Mahdi, Ghufran Alshareedah, and Maytham Safar. *Mobile Communication Network*. 01 2012.

[22] Muhammad Farooq, Muhammad Ishtiaq Ahmed, and Usman M Al. Future generations of mobile communication networks. *Academy of Contemporary Research Journal*, 2(1):24–30, 2013.

[23] Mudit Ratana Bhalla, Anand Vardhan Bhalla, et al. Generations of mobile wireless technology: A survey. *International Journal of Computer Applications*, 5(4):26–32, 2010.

[24] Stephen Carroll. *Public Switched Telephone Network Definition - ULTATEL*, Jan 2024. https://www.ultatel.com/terms/communication-platforms/public-switched-telephone-network/.

[25] Marie Look. *Millions of Phones Use GSM ... Meaning What, Exactly?*, Apr 2024. https://electronics.howstuffworks.com/question537.htm.

[26] Chiradeep BasuMallick. *What Is GSM (Global System for Mobile Communications)? Meaning, Working, Architecture, and Applications*, Dec 2022. https://www.spiceworks.com/tech/networking/articles/what-is-gsm/.

[27] Anshuman Singh. *Difference Between Analog and Digital Signal*, May 2024. https://www.shiksha.com/online-courses/articles/difference-between-analog-and-digital-signals/.

[28] Wesley Chai and Corinne Bernstein. *3GPP (3rd Generation Partnership Project)*, mar 2024. https://www.techtarget.com/searchnetworking/definition/3rd-Generation-Partnership-Project-3GPP.

[29] Sabine Dahmen-Lhuissier. *4th Generation (LTE)*. https://www.etsi.org/technologies/mobile/4G.

[30] Communication Systems. *What is the S1 interface in the LTE network architecture?*, 2024. https://www.linkedin.com/advice/3/what-s1-interface-lte-network-architecture.

[31] Konstantinos Alexandris, Navid Nikaein, R. Knopp, and Christian Bonnet. Analyzing x2 handover in lte/lte-a. 05 2016.

[32] Ramraj Dangi, Gaurav Choudhary, Nicola Dragoni, Praveen Lalwani, Utkarsh Khare, and Souradeep Kundu. 6g mobile networks: Key technologies, directions, and advances. In *Telecom*, volume 4, pages 836–876. MDPI, 2023.

[33] Cong T Nguyen, Yuris Mulya Saputra, Nguyen Van Huynh, Tan N Nguyen, Dinh Thai Hoang, Diep N Nguyen, Van-Quan Pham, Miroslav Voznak, Symeon Chatzinotas, and Dinh-Hieu Tran. Emerging technologies for 6g non-terrestrial-networks: From academia to industrial applications. *arXiv preprint arXiv:2403.07763*, 2024.

[34] Xiangming Zhu and Chunxiao Jiang. Integrated satellite-terrestrial networks toward 6g: Architectures, applications, and challenges. *IEEE Internet of Things Journal*, 9(1):437–461, 2021.

[35] Catherine. *What is a base station?*, 2024. https://www.fibermall.com/blog/what-is-a-base-station.htm.

[36] Salah Bendary. *Cellular Network Infrastructure: From Antenna to BBU*, May 2024. https://www.linkedin.com/pulse/cellular-network-infrastructure-from-antenna-bbu-salah-bendary-x9dof/.

[37] Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, et al. A survey of large language models. *arXiv preprint arXiv:2303.18223*, 2023.

[38] Ajitesh Kumar. *Large Language Models (LLMs): Types, Examples*, Jan2024. https://vitalflux.com/large-language-models-concepts-examples/.

[39] CellStrat. *Real-World Use Cases for Large Language Models (LLMs)*, Apr 2023. =https://cellstrat.medium.com/real-world-use-cases-for-large-language-models-llms-d71c3a577bf2.

[40] R Zhao, Liu Ying, Congjian Deng, et al. Design of a large language model for improving customer service in telecom operators. 2024.

[41] Tasnim Ahmed, Nicola Piovesan, Antonio De Domenico, and Salimur Choudhury. Linguistic intelligence in large language models for telecommunications. *arXiv preprint arXiv:2402.15818*, 2024.

[42] Hang Zou, Qiyang Zhao, Lina Bariah, Yu Tian, Mehdi Bennis, Samson Lasaulce, Merouane Debbah, and Faouzi Bader. Genainet: Enabling wireless collective intelligence via knowledge transfer and reasoning. *arXiv preprint arXiv:2402.16631*, 2024.

[43] Nicola Piovesan, Antonio De Domenico, and Fadhel Ayed. Telecom language models: Must they be large? *arXiv preprint arXiv:2403.04666*, 2024.

[44] Lina Bariah, Qiyang Zhao, Hang Zou, Yu Tian, Faouzi Bader, and Merouane Debbah. Large generative ai models for telecom: The next big thing? *IEEE Communications Magazine*, 2024.

[45] Athanasios Karapantelakis, Mukesh Shakur, Alexandros Nikou, Farnaz Moradi, Christian Orlog, Fitsum Gaim, Henrik Holm, Doumitrou Daniil Nimara, and Vincent Huang. Using large language models to understand telecom standards. *arXiv preprint arXiv:2404.02929*, 2024.

[46] Mohamad Nabeel, Doumitrou Daniil Nimara, and Tahar Zanouda. Test code generation for telecom software systems using two-stage generative model. *arXiv preprint arXiv:2404.09249*, 2024.

[47] Ali Maatouk, Fadhel Ayed, Nicola Piovesan, Antonio De Domenico, Merouane Debbah, and Zhi-Quan Luo. Teleqna: A benchmark dataset to assess large language models telecommunications knowledge. *arXiv preprint arXiv:2310.15051*, 2023.

[48] Muhammad Usman Hadi, Rizwan Qureshi, Abbas Shah, Muhammad Irfan, Anas Zafar, Muhammad Bilal Shaikh, Naveed Akhtar, Jia Wu, Seyedali Mirjalili, et al. A survey on large language models: Applications, challenges, limitations, and practical usage. *Authorea Preprints*, 2023.

[49] Lina Bariah, Qiyang Zhao, Hang Zou, Yu Tian, Faouzi Bader, and Merouane Debbah. Large generative ai models for telecom: The next big thing? *IEEE Communications Magazine*, 2024.

[50] LLM team. *Large Language Models Explained*. https://www.nvidia.com/en-us/glossary/large-language-models/, year=.

[51] Ali Maatouk, Nicola Piovesan, Fadhel Ayed, Antonio De Domenico, and Merouane Debbah. Large language models for telecom: Forthcoming impact on the industry. *arXiv preprint arXiv:2308.06013*, 2023.

[52] Sanjeev kapoor. Large language models the basics you need to know, Sep 2023.

[53] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[54] Debbah Merouane. *Large Language Models for the Telecom: The Next Big Thing*, Dec 2023. https://ieeetv.ieee.org/channels/ieee-future-networks/large-language-models-for-the-telecom-the-next-big-thing-merouane-debbah?check_logged_in=1.

[55] LLMs team. *Large Language Models (LLMs)*, Oct 2023. https://www.databricks.com/glossary/large-language-models-llm.

[56] Jean Kaddour, Joshua Harris, Maximilian Mozes, Herbie Bradley, Roberta Raileanu, and Robert McHardy. Challenges and applications of large language models. *arXiv preprint arXiv:2307.10169*, 2023.

[57] Henrik Holm. *Bidirectional encoder representations from transformers (bert) for question answering in the telecom domain.: Adapting a bert-like language model to the telecom domain using the electra pre-training approach*, 2021.

[58] Gupta Gaurav. *Is Python the best fit for artificial intelligence in web development?*, Apr 2024. https://squareboat.com/blog/Is-Python-the-best-fit-for-artificial-intelligence-in-web-development.

[59] GeeksforGeeks. *How to use Google Colab*, Jun 2024. https://www.geeksforgeeks.org/how-to-use-google-colab/, journal=GeeksforGeeks.

[60] Youdi Gong, Guangzhen Liu, Yunzhi Xue, Rui Li, and Lingzhong Meng. A survey on dataset quality in machine learning. *Information and Software Technology*, page 107268, 2023.

[61] Mingyang Wan, Daochen Zha, Ninghao Liu, and Na Zou. Modeling techniques for machine learning fairness: A survey. *arXiv preprint arXiv:2111.03015*, 2021.

[62] Lei Zhang, Shuai Wang, and Bing Liu. Deep learning for sentiment analysis: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4):e1253, 2018.

[63] Vik Paruchuri. *43 Free Datasets for Building an Irresistible Portfolio (2024)*, Mar 2024. https://www.dataquest.io/blog/free-datasets-for-projects/.

[64] Alex Freberg. *Best Places to Find Datasets for Your Projects*, Jun 2024. https://www.youtube.com/watch?v=PExdWWcxmro, journal=YouTube.