

ITUEvents

Workshop

At the crossroads of standards and research: AI/ML datasets for future networks

16 July 2024

Geneva, Switzerland

itu.int/go/Standards_Research

**TeleQnA: A Benchmark Dataset to
Assess Large Language Models
Telecommunications Knowledge**

Antonio De Domenico
Huawei Technologies



Antonio De Domenico

- **About**

- Senior Researcher at Huawei Technologies, France
- Active in ITU-T SG 13- Future networks and emerging network technologies
- Leading IEEE ETI GenAI Dataset and Competitions activities
- Co-leading network energy efficiency in NGMN Green Future Network project

- **Current Research**

- Integrating AI/ML in 5G and beyond RANs

- **Contact**

- antonio.de.domenico@huawei.com



The Race of Language Models

Some key observations:

Doubling time of **LLM compute requirements**: 4-6 months

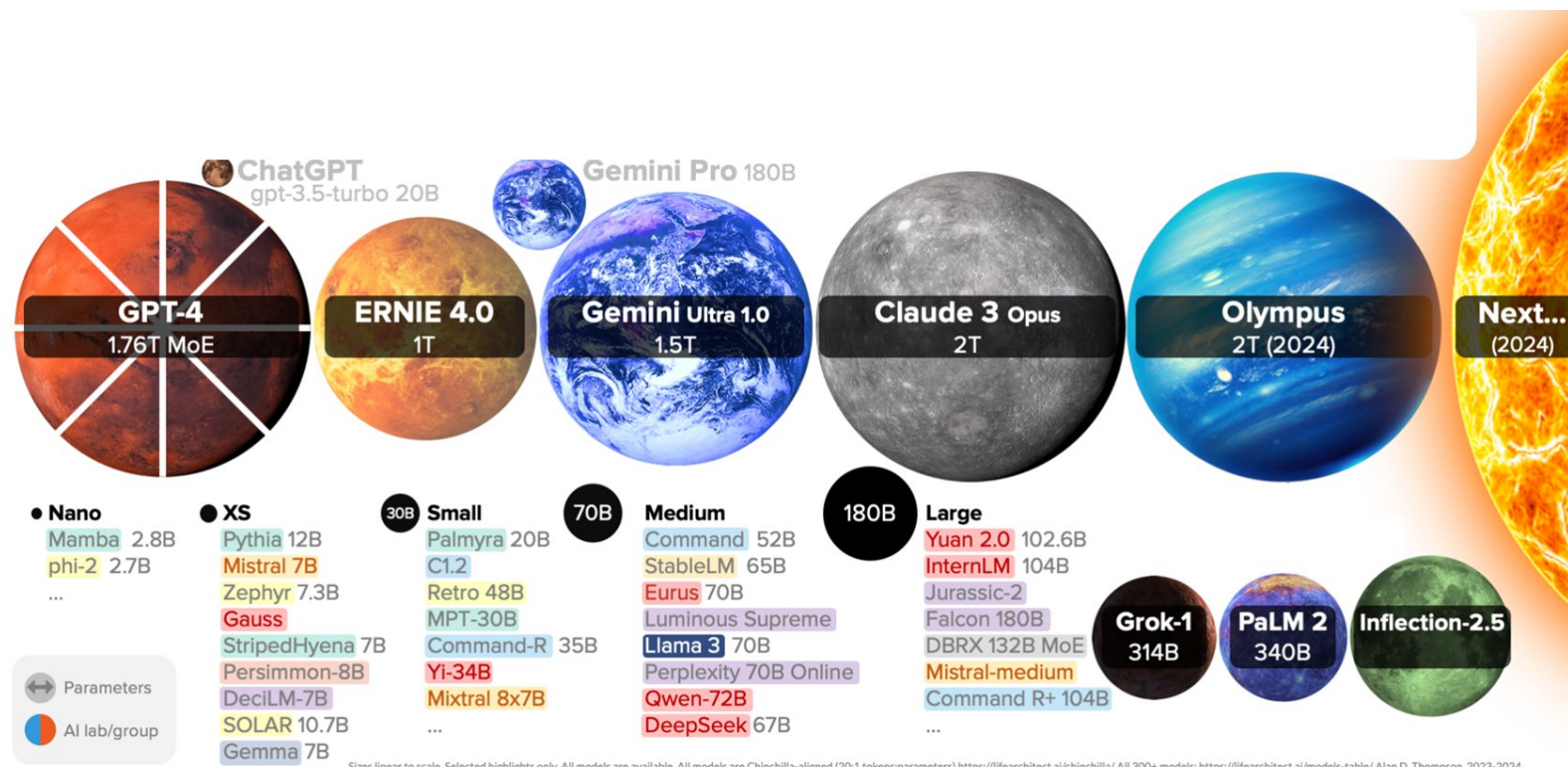
- Way beyond Moore's law

Training cost of 10B model~1M\$

Open source models' performance lags private models' one by ~1 year

Chinchilla scaling laws: 20 text token/parameters

- 1.4T token for 70B model

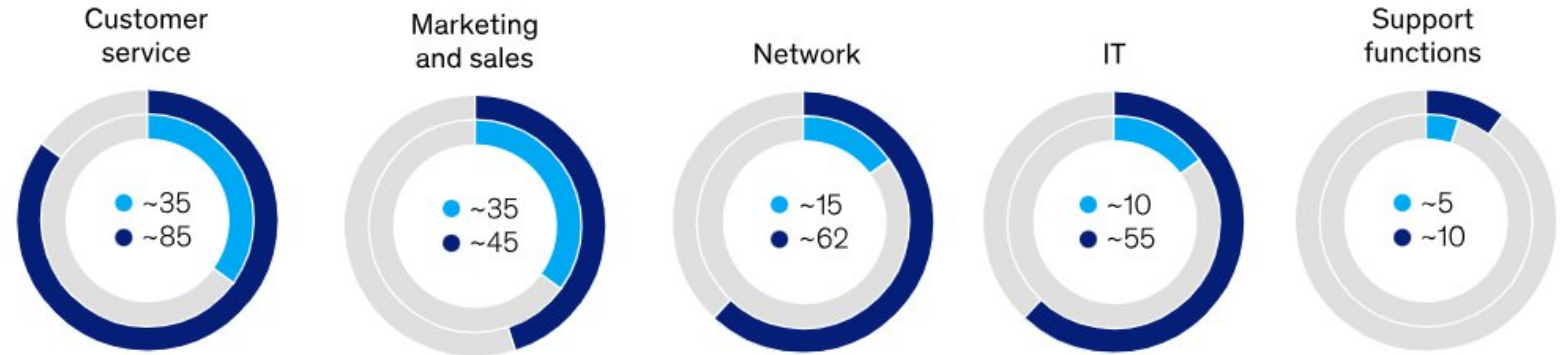


LLMs Use Cases and Expected Value in Telecoms

- Customer service AI **chatbots**
- Create **personalized** messages and visual media
- **Analyzing** specifications and maintenance **data**
- Streamline the entire **software life cycle**
- Improve employee **productivity**

Distribution of impact by business domain¹

● Share of total impact, % ● Share of surveyed business leaders focused on domain, %



Example use cases

Customer-facing chatbots, call-routing performance, agent copilots, bespoke invoice creation

Content generation, hyperpersonalization, copilots for store personnel, customer sentiment analysis and synthesis

Network inventory mapping, network optimization via customer sentiment analysis, enabling self-healing via customer sentiment analysis on network problems

Copilots for software development, synthetic data generation, code migration, IT support chatbots

Procurement optimization, workplace productivity, internal knowledge management, content generation, HR Q&A

¹The distribution of impact by business domain is based on our experience working with telco companies to deploy gen AI and includes impacts on both capital expenditure and EBITDA.

Forthcoming Impact on the Telecoms Industry

ARTIFICIAL INTELLIGENCE AND DATA SCIENCE FOR COMMUNICATIONS

Large Language Models for Telecom: Forthcoming Impact on the Industry

Ali Maatouk, Nicola Piovesan, Fadhel Ayed, Antonio De Domenico, and Merouane Debbah

The authors delve into the inner workings of large language models, providing insights into their current capabilities and

ABSTRACT

Large language models (LLMs) – AI-driven models that can achieve general-purpose language understanding and generation – have emerged as a transformative force, revolutionizing

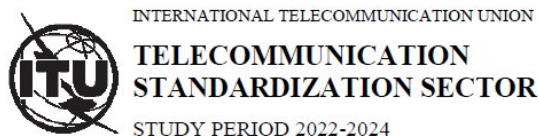
diverse domains [3], surpassing traditional NLP applications like machine translation and sentiment analysis. In fact, through domain-specific data, they can excel in tasks related to that particular domain. For instance, in medicine, LLMs may play a crucial role in encoding clinical knowledge

Our goal: use LLMs to pave the way towards Network Automation by bridging the gaps between costumers, researchers, and engineers.

we delve into the inner workings of LLMs, provid-

based virtual tutor and classroom assistant.

ITU-T SG 13: Telecoms Requirements and GenAI Capabilities



SG13-TD763/WP1
STUDY GROUP 13
Original: English

Question(s): 20/13

Geneva, 4-15 March 2023

TD

Source:	Editors	
Title:	Draft new Technical Report ITU-T TR-GenAI-Telecom Networks: “Requirements and methodology for deploying and assessing Generative AI models in telecom networks”	
Contact:	Antonio De Domenico Huawei Technologies Düsseldorf GmbH Germany	Tel: +33-673462921 E-mail: antonio.de.domenico@huawei.com
Contact:	Nicola Piovesan, Huawei Technologies Co. Ltd. China	Tel: +33-637153889 E-mail: nicola.piovesan@huawei.com
Contact:	Xiaoqing Xu, China Telecom China	Tel: +86-18998490768 E-mail: xuxq@chinatelecom.cn
Contact:	Yongsheng Liu China Unicom China	Tel: +86-18601106253 E-mail: liuys170@chinaunicom.cn
Contact:	Qiyang Zhao Technology Innovation Institute UAE	Tel: +971-509663216 Email: Qiyang.Zhao@tii.ae



CORRESPONDENCE GROUP ON
DATASETS (CGDATASETS)

CGD-O-01

Original: English

Question(s): N/A

cgdatasets meeting, Virtual, 12 April 2024

OUTPUT DOCUMENT

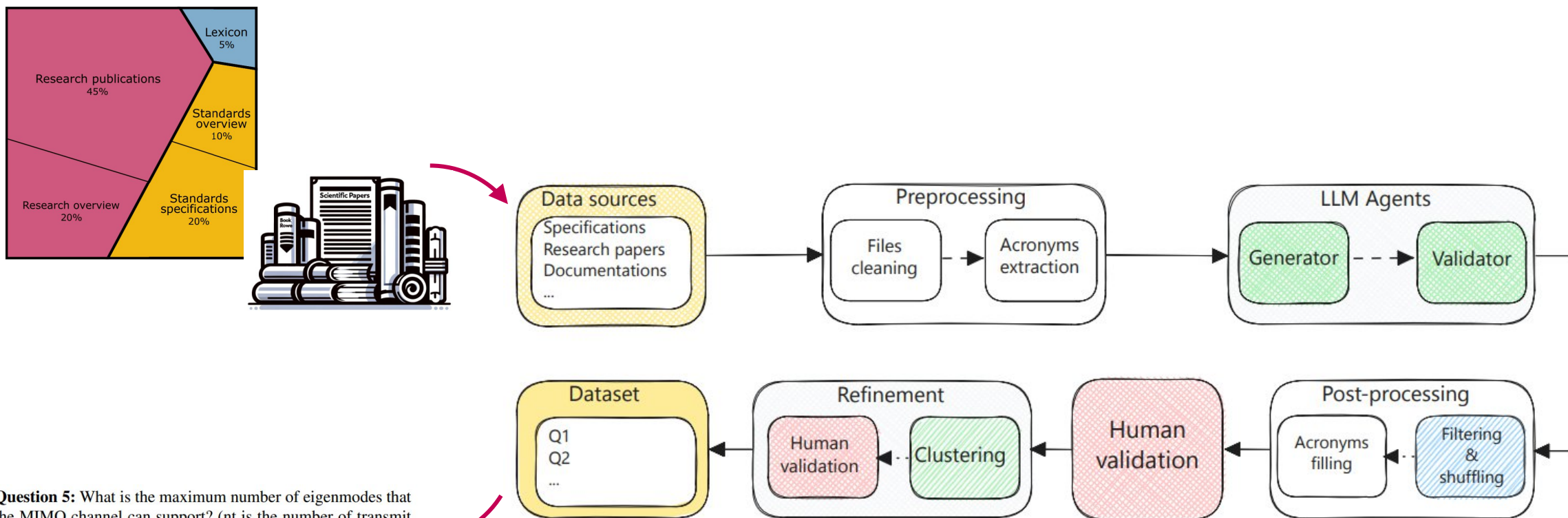
Source:	Editors	
Title:	Draft Technical Report “Datasets standardization approaches for datasets applicable for AI/ML in networks” – deliverable of the Correspondence Group on Datasets (cgdatasets)	
Contact:	Vishnu Ram O.V Independent Expert	Email: vishnu.n@ieee.org
Keywords:	Artificial Intelligence, datasets, Machine Learning, networks	
Abstract:	This document is the updated version of the Technical Report on “Datasets standardization approaches for datasets applicable for AI/ML in networks”, deliverable of the Correspondence Group on Datasets (cgdatasets).	

This document is the updated version of the Technical Report on “Datasets standardization approaches for datasets applicable for AI/ML in networks”, deliverable of the Correspondence Group on Datasets (cgdatasets) – output of the 12 April 2024 cgdatasets meeting (it includes some post-12 April meeting editorial clean-up).

- ITU-T SG13, TR-GenAI-Telecom, “Requirements and methodology for deploying and assessing Generative AI models in telecom networks” https://www.itu.int/itu-t/workprog/wp_item.aspx?isn=19366
- Correspondence Group for datasets applicable for AI/ML in networks (CG-datasets for AI/ML in networks) <https://extranet.itu.int/sites/itu-t/studygroups/2022-2024/sg13/cg-datasets/SitePages/Home.aspx>



Assessing the Telecoms Knowledge of LLMs



Question 5: What is the maximum number of eigenmodes that the MIMO channel can support? (n_t is the number of transmit antennas, n_r is the number of receive antennas)

- Option 1: n_t
- Option 2: n_r
- Option 3: $\min(n_t, n_r)$
- Option 4: $\max(n_t, n_r)$

Answer: Option 3: $\min(n_t, n_r)$

Explanation: The maximum number of eigenmodes that the MIMO channel can support is $\min(n_t, n_r)$.

Category: Research publications

Automatic evaluation based on a LLM-generated dataset with human in the loop!!!

Assessing the Telecoms Knowledge of LLMs

Submitted to IEEE Wireless Communications

TeleQnA: A Benchmark Dataset to Assess Large Language Models Telecommunications Knowledge

Ali Maatouk^{*†}, Fadhel Ayed^{*†}, Nicola Piovesan[†], Antonio De Domenico[†], Merouane Debbah[‡], Zhi-Quan Luo[§]

[†]Paris Research Center, Huawei Technologies, Boulogne-Billancourt, France

[‡]Khalifa University of Science and Technology, Abu Dhabi, UAE

[§]The Chinese University of Hong Kong, Shenzhen, China

Abstract—We introduce TeleQnA¹, the first benchmark dataset designed to evaluate the knowledge of Large Language Models (LLMs) in telecommunications. Comprising 10,000 questions and answers, this dataset draws from diverse sources, including standards and research articles. This paper outlines the automated question generation framework responsible for creating this dataset, along with how human input was integrated at various stages to ensure the quality of the questions. Afterwards, using the provided dataset, an evaluation is conducted to assess the capabilities of LLMs, including GPT-3.5 and GPT-4. The results highlight that these models struggle with complex standards-related questions but exhibit proficiency in addressing general

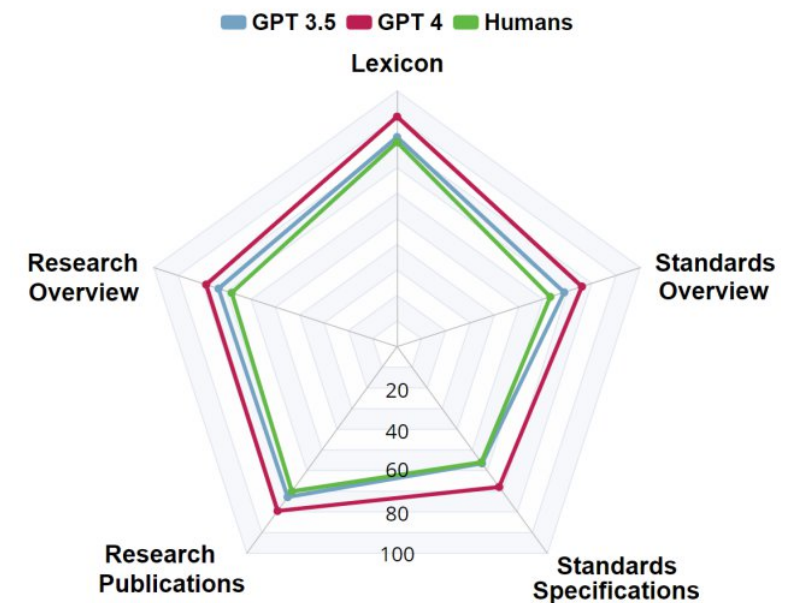
be observed in other domains, such as medicine and finance, where benchmark datasets like MultiMedQA [1] and FLUE [8] have been introduced to assess the proficiency of LLMs in these fields.

As LLMs find their way into the telecommunications industry, a clear and pressing issue arises—there is a notable absence of a benchmark dataset designed to evaluate these models' proficiency in telecom. Consequently, there is an urgent need for such a dataset, as highlighted in various prior research (e.g., [9]). This paper aims to bridge this gap

The Telecom Knowledge of GPTs

- **GPT-4 consistently outperforms GPT-3.5**, demonstrating around **7% improvement** across all categories
- LLMs exhibit **exceptional performance in the lexicon category**, which encompasses general telecom knowledge and terminology
- LLMs face challenges when confronted with more intricate questions related to **standards**, with the highest performing model, GPT-4, achieving a modest 64% accuracy in this domain
- LLMs and active professionals exhibit comparable performance in general telecom knowledge.

	Mistral 7b	Mixtral (MoE)	GPT-3.5	GPT-4	Humans
Lexicon (500)	56.80	83.80	82.20	86.80	80.33
Research overview (2000)	51.60	70.70	68.50	76.25	63.66
Research publications (4500)	49.27	70.20	70.42	77.62	68.33
Standards overview (1000)	46.20	66.73	64.00	74.40	61.66
Standards specifications (2000)	35.60	55.85	56.97	64.78	56.33
Overall accuracy (10000)	47.07	67.74	67.29	74.91	64.86

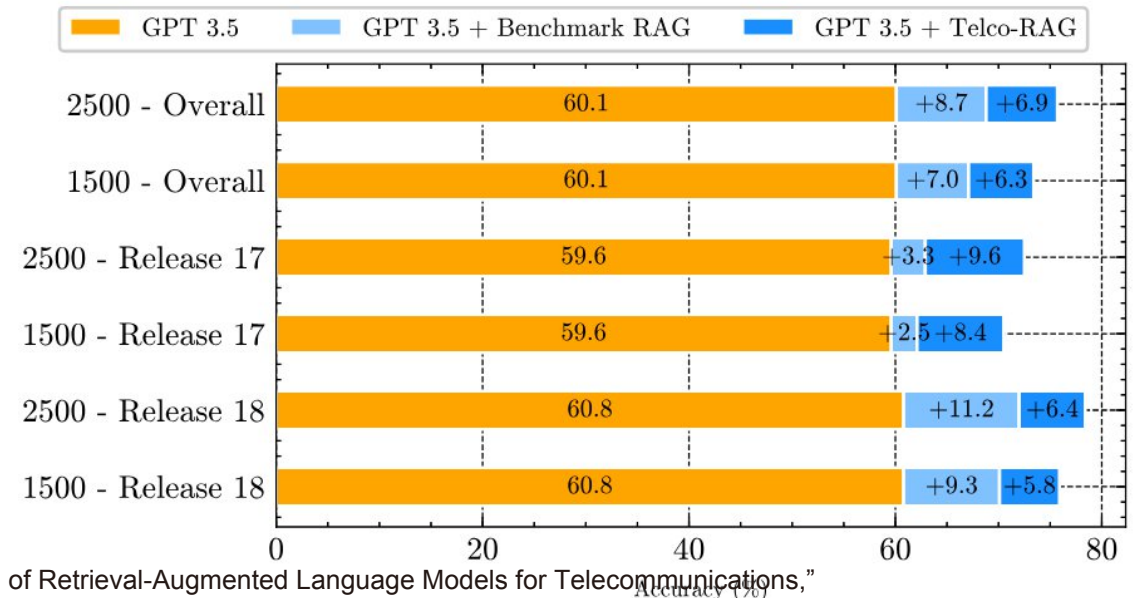
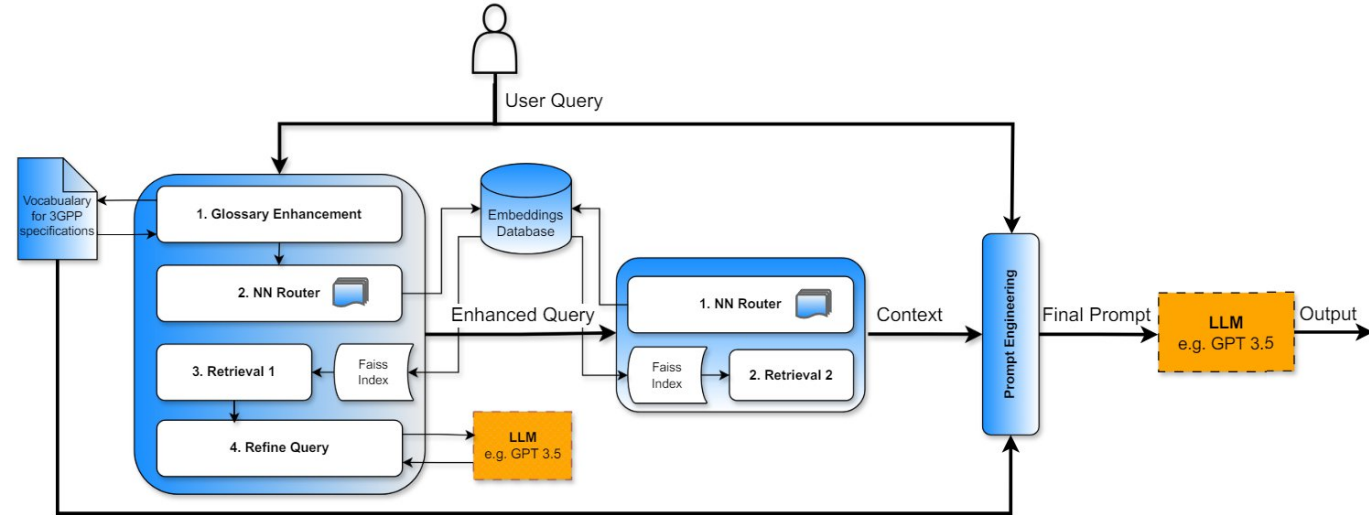


We need LMs specialized on Telecoms Networks

- A. Maatouk, F. Ayed, N. Piovesan, A. De Domenico, M. Debbah, Zhi-Quan Luo, "TeleQnA: A Benchmark Dataset to Assess Large Language Models Telecommunications Knowledge", *arXiv preprint arXiv:2310.15051* <https://arxiv.org/abs/2310.15051>
- <https://github.com/netop-team/TeleQnA>

Telco-RAG: A RAG Framework to Handle Queries on Standards

- E2E design for improving LLM capabilities on Telcos standard questions
 - Hyperparameters Optimization
 - Query Augmentation
 - Enhancing the RAM Usage
 - Prompt Engineering
- Telco-RAG leads to an **average improvement of 6.6% and 14.45% compared to GPT 3.5** with and without the Benchmark RAG



Telco-RAG: A RAG Framework to Handle Queries on Standards



- AL Bornea, F Ayed, A De Domenico, N Piovesan, A Maatouk, "Telco-RAG: Navigating the Challenges of Retrieval-Augmented Language Models for Telecommunications," arXiv preprint arXiv:2404.15939, <https://arxiv.org/pdf/2404.15939>
- <https://github.com/netop-team/Telco-RAG>

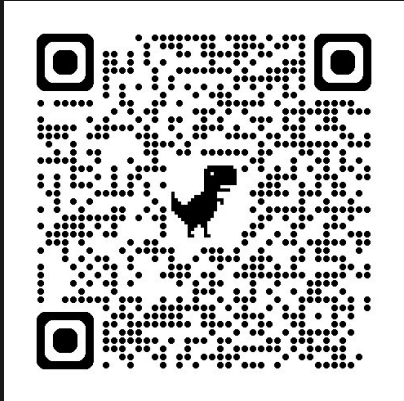
ITUEvents



AI for Good

Machine Learning in 5G Challenge

*Specializing Large Language
Models for Telecom Networks*



aiforgood.itu.int



Conclusion

LLMs and GenAI will have a large impact on future networks, both from the **requirements** and **capabilities** perspectives

The telco ecosystem needs to identify the most relevant **use cases** for LLMs in telcos, and the **required functionalities/capabilities** proper to telcos (beyond standard NLP)

- Specialized telco models: Telecom knowledge, coding, reasoning, calculus, etc

Based on these, the industry can jointly define proper **evaluation methodologies**:

- **Tests, datasets, benchmarking metrics, and platforms**

This is just the beginning of the journey

Thank you.



antonio.de.domenico@huawei.com