AI for Good Discovery

# AI/ML in 5G

## *AI Telco Troubleshooting Challenge global launch*

Robert Gazda, Antonio De Domenico, and Dario Sabella

**21 November 2025**

50+ UN PARTNERS

ITU

# Using AI Models for Root Cause Analysis in Telecom Networks: The Problem

- *Reducing the cost of network faults due to hardware failures to software miss-configurations is a key challenge for service providers.*
- Using Telelogs researchers have shown that specialized LLMs can support Root Cause Analysis (see next slide)

- Open Research questions:
  - How to build models that generalize to multiple scenarios, e.g., unseen faults, data, networks?
  - Can small language models, run on edge servers, efficiently perform this task?
- Optional research challenges
  - How security is addressed, in this edge-cloud continuum, also to protect critical data?

**https://aiforgood.itu.int/ai-telco-troubleshooting-challenge/**

In partnership with the world's leading community organizations: **GSMA, ETSI, IEEE, ITU, and TM Forum**
Supported by headline and technology partners including: **Huawei, InterDigital, NextGCloud, RelationalAI, and xFlow Research**

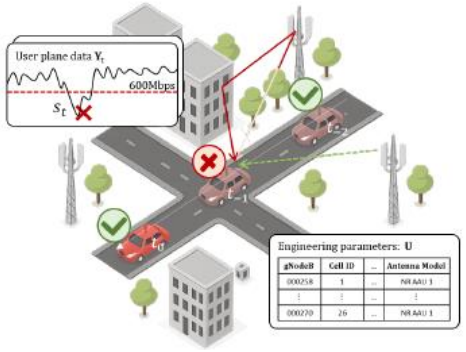
# Reasoning Language Models for Root Cause Analysis in 5G Wireless Networks: Background results

- *Reducing the cost of network faults due to hardware failures to software miss-configurations is a key challenge for service providers.*

**TeleLogs includes training and test datasets** [1] to

1. Evaluate if LLMs can support engineers in **Root Cause Analysis**
2. Specialize LLMs in Root Cause Analysis

| Ref. Model | Accuracy (%) | | | |
|---|---|---|---|---|
| | Test dataset | | Randomized dataset | |
| | Pass@1 | Maj@4 | Pass@1 | Maj@4 |
| **Base** | | | | |
| Qwen2.5 1.5B-Instruct | 11.25 | 11.6 | 9.15 | 9.8 |
| Qwen2.5 7B-Instruct | 12.05 | 10.8 | 11.5 | 11.8 |
| Qwen2.5 32B-Instruct | 18.85 | 19.6 | 18.05 | 18.7 |
| **Reasoning** | | | | |
| DeepSeek-R1 Distill-Llama-70B | 29.42 | 34.84 | 29 | 32.18 |
| QwQ 32B | 33.62 | 39 | 32.14 | 38.86 |
| Qwen3 32B | 33.77 | 37.04 | 31.37 | 36.23 |
| **Proposed** | | | | |
| Qwen2.5-RCA 1.5B | 87.56 | 87.73 | 75.90 | 77.08 |
| Qwen2.5-RCA 7B | 87.01 | 88.89 | 77.95 | 80.32 |
| Qwen2.5-RCA 32B | **95.86** | **96.18** | **93.23** | **95.02** |




Root cause analysis scenario

**Prompt template $\mathcal{T}$**

Analyse the user plane data and signalling plane data of a 5G wireless network in the following drive test, as well as the site engineering parameters (tilt, etc.) and configuration (handover parameters, etc.).

Analyse the reasons for $s_t$. Choose the most likely root cause from the following $k$ reasons and include the root cause number in the final answer, enclosed in \boxed{}.

[Here are the possible reasons **C**]

[Here are the network configuration data **U**]

[Here are the signalling plane data $\mathbf{Y}_t$]

Prompt template in TeleLogs[1]

Using Telelogs, researchers [2] demonstrated that even **small LLMs can perform greatly when specialized using SFT+RL**

However, **AI models generalization capabilities are** still **limited**

- A model trained on a dataset will not perform as good even in a slightly different scenario (different data, network issues, labels)

[1] https://huggingface.co/datasets/netop/TeleLogs
[2] Reasoning Language Models for Root Cause Analysis in 5G Wireless Networks, subm. to IEEE INFOCOM, https://arxiv.org/pdf/2507.21974

# AI Telco Troubleshooting Challenge: Tracks

## Tracks

**Develop novel solutions or combine existing methods such as supervised fine-tuning and reinforcement learning to:**

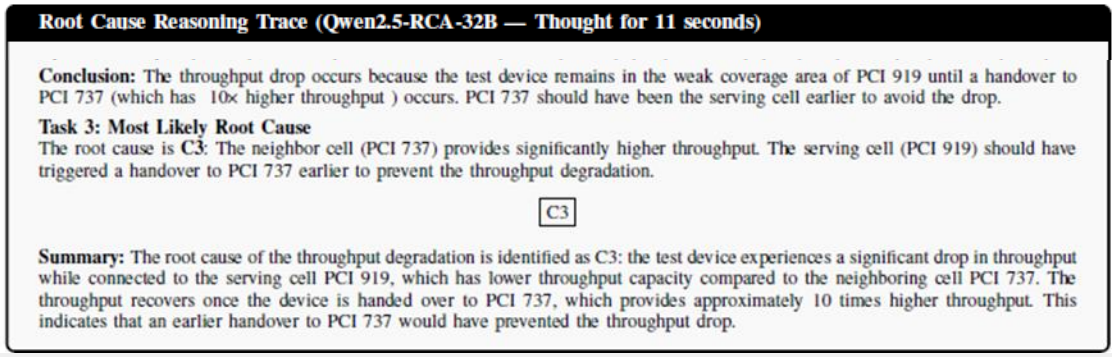
1. **Specialize a reasoning LLM on Telecom Network Troubleshooting**
   Enhance the accuracy of Qwen3-32B when answering to questions in the TeleLogs dataset.

2. **Specialize a mid-cloud LLM on Telecom Network Troubleshooting**
   Enhance the accuracy of Qwen2.5-7B-Instruct when answering to questions in the TeleLogs dataset.

3. **Specialize an edge-cloud LLM on Telecom Network Troubleshooting**
   Enhance the accuracy of Qwen2.5-1.5B-Instruct when answering to questions in the TeleLogs dataset.

## Challenges:

1. **Complexity and diversity of the questions in the dataset**
2. **LLM Hallucinations and Fabrications**
3. **Limited explainability**



**Root Cause Reasoning Trace (Qwen2.5-RCA-32B — Thought for 11 seconds)**

**Conclusion:** The throughput drop occurs because the test device remains in the weak coverage area of PCI 919 until a handover to PCI 737 (which has 10× higher throughput ) occurs. PCI 737 should have been the serving cell earlier to avoid the drop.

**Task 3: Most Likely Root Cause**
The root cause is **C3**: The neighbor cell (PCI 737) provides significantly higher throughput. The serving cell (PCI 919) should have triggered a handover to PCI 737 earlier to prevent the throughput degradation.

C3

**Summary:** The root cause of the throughput degradation is identified as C3: the test device experiences a significant drop in throughput while connected to the serving cell PCI 919, which has lower throughput capacity compared to the neighboring cell PCI 737. The throughput recovers once the device is handed over to PCI 737, which provides approximately 10 times higher throughput. This indicates that an earlier handover to PCI 737 would have prevented the throughput drop.

## Competition platform:

This competition uses Zindi as a platform for discussion/support, data retrieval, submission and automatic evaluation

https://zindi.africa/competitions/the-ai-telco-troubleshooting-challenge

# AI Telco Troubleshooting Challenge: Evaluation Criteria

1 **Accuracy on the private dataset**, which include troubleshooting problems and general purpose questions to evaluate model knowledge retention

The participants will work on the public dataset for building and testing their model.

- The accuracy of the submission will be reported on the leaderboard and based on the test set in the public dataset

They will receive towards the end of the competition a private dataset, which does not include the correct answers to the associated questions

- The goal of this dataset is to 1) avoid submission overfitting and 2) test the submitted model knowledge retention
- The accuracy of the submission on the private set will be also include on the leaderboard during the last phase of the competition and used for the final evaluation

**Accuracy** of the submitted model will be computed using the **Pass@1** metric denotes the fraction of correct replies to (*N*) questions with *4* response generations for each answer,

$$A = \frac{1}{NK}\sum_{i=1}^{N}\sum_{j=1}^{K}\delta\left(y_i - \widehat{y_{i,j}}\right)$$

Where $\delta$ indicates if the ith ground-truth $y_i$ and the kth response generation to the jth question $\widehat{y_{i,j}}$ are the same

# AI Telco Troubleshooting Challenge: Evaluation Criteria

2 **Reproducibility** of specialized model and associated code

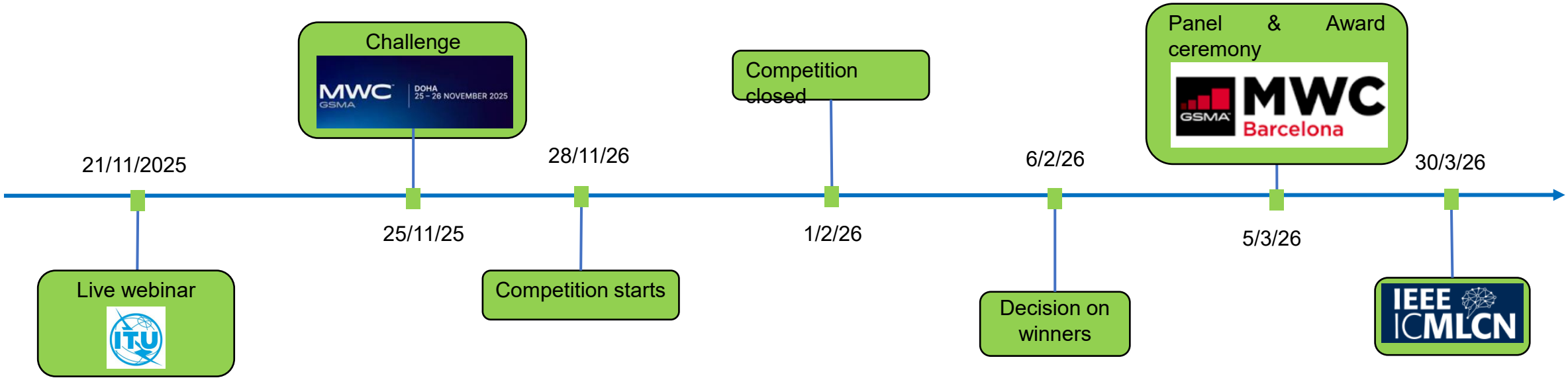
To foster results reproducibility and ensure the submission quality, we require the participants to publicly share the model obtained and the associated code

3 **Explainability** of the work based on the delivered short **report** including the analysis on security and privacy aspects

In this report, in addition to **explaining their work** and **providing a description of the steps to obtain the achieved results**, participants are strongly invited to include a **short text on one or more of the following aspects**:

- Data privacy and compliance;
- Model security risks;
- Data and model access control and transparency;
- Edge computing considerations and vulnerabilities;
- Data governance issues.

# AI Telco Troubleshooting Challenge: Overall Timeline




**The participants will be invited to submit paper to a workshop organized at IEEE ICMLCN 2026 (https://icmlcn2026.ieee-icmlcn.org/) to further disseminate their work and results to the academic community**

# AI Telco Troubleshooting Challenge: Prizes

35 k€ shared equally for the team winners of each of the 3 challenge tracks

- Trip and stay in Barcelona for a representative of each team winner
- MWC Barcelona 2026 ticket for a representative of each team winner
- You will add your team results to the GSMA Open-Telco LLM Benchmarks

# Questions?

AI for Good
Discovery

## AI/ML in 5G

*AI Telco Troubleshooting Challenge global launch*

**21 November 2025**

50+ UN PARTNERS

ITU