Poster: Could Large Language Models Perform Network Management?

Zine el abidine Kherroubi, Monika Prakash, Jean-Pierre Giacalone and Michael Baddeley*
{zine.kherroubi,monika.prakash,jean-pierre.giacalone,michael.baddeley}@tii.ae
Technology Innovation Institute
Abu Dhabi, UAE

ABSTRACT

Modern wireless communication systems have become increasingly complex due to the proliferation of wireless devices, increasing performance standards, and growing security threats. Managing these networks is becoming more challenging, requiring the use of advanced network management methods and tools. AI-driven network management systems such as Self-Optimizing Networks (SONs) are gaining attention. On the other hand, Large Language Models (LLMs) have been demonstrating exceptional zero-shot learning and generalization capabilities across several domains. In this paper, we leverage the potential of LLMs with SONs to enhance future network management systems. Specifically, we benchmark the use of various LLMs such as GPT-4, Llama, and Falcon, in a zero-shot setting based on their real-time network configuration recommendations. Our results indicate promising prospects for integrating LLMs into future network management systems.

CCS CONCEPTS

• **Networks → Network management**.

KEYWORDS

LLM, Wireless, GPT, Falcon, Network Management, Zero-Shot

1 INTRODUCTION

Real-time network management has become a necessity due to the increased complexity, scale and diversity of modern wireless communication systems. The integration of 5G, IoT, AI, edge computing, and SDN/NFV has made real-time decision-making and dynamic adjustments critical to maintaining the performance, security, and efficiency of wireless networks. Most of the existing network management methods are often reactive addressing problems after their occurrence. However, with recent advancements in AI and Machine Learning (ML), proactive and predictive network management techniques such as Self Optimizing Networks (SONs) are gaining attention [1, 2].

Recently, Large Language Models (LLMs) have demonstrated exceptional reasoning and generalization capabilities, which open doors to widespread applications beyond natural language processing tasks. LLMs are now being used in diverse fields such as energy [3], finance [4], and transportation [5]. In [6], a telecom-specific LLM, TelecomGPT, was fine-tuned to perform different tasks, such as mathematical modeling and content analysis in the telecoms. An LLM-based Intent translation system [7] was also proposed to allows users to express Intents in natural language, and subsequently converts them into Network Service Descriptors (NDSs). To demonstrate how LLMs can simplify and automate complex network management tasks, authors of [8] developed a model-agnostic network configuration benchmark for LLMs called NetConfEval. While these works emphasis the promising perspectives of using LLMs in wireless communication, they remain largely focused on task of natural language understanding and processing. We propose that the real opportunity lies in expanding the role of LLMs as autonomous decision-makers and planners for network management, thus offering a more adaptive, proactive approach to managing modern wireless networks. We therefore propose a benchmark study where we will use and compare various pre-trained LLMs, and instruct them as a zero-shot learners to provide network configuration recommendations.

2 SYSTEM MODEL

We address a scenario of a mission critical network that supports essential services like smart healthcare, smart manufacturing and first responders – where real-time network management is crucial due to the high stakes involved in maintaining continuous, secure, and reliable communication. Specifically, Wi-Fi based infrastructure-less mesh networks are considered. As shown in Fig. 1, each node in the mesh network shares periodic network status reports, which includes metrics such as TX/RX throughput, latency, packet loss, and neighbor nodes. Additionally, they report events like channel interference and jamming detection, which require immediate action to be taken. Based on the observed network states, the LLM is instructed to perform zero-shot reasoning for network configuration recommendations, as illustrated in Table 1. To do this, we establish the context for the LLM to act as *an expert in network management*, as defined by *the system prompt*. In addition, *the user prompt* provides network state observations, a list of valid actions, and a step-by-step task description, as also illustrated in Table 1. The LLM then generates a response by selecting an action, which is marked with the tag *<ACTION>chosen action by the LLM</ACTION>*.

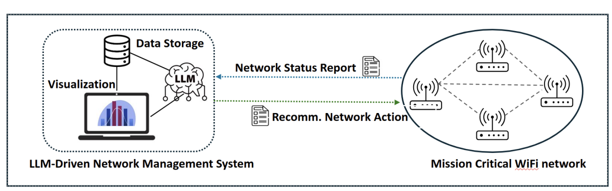


Figure 1: System Model

3 BENCHMARKING RESULTS

To assess the performance of LLMs for network management, we prompted various models and compared the quality of their generated responses. Our analysis included state-of-the-art open-source models such as the Llama and Falcon series, as well as GPT models accessed via OpenAI APIs. The evaluation was based on three

Table 1: Prompt template

System prompt

You are a network monitoring expert, and you monitor a wireless mesh network. When there is a network security threat such as malicious traffic, jamming, etc., you need to take a valid action among the valid actions set to mitigate it. Sometimes, it will be a network performance related update. For example, when best neighbors of a node is received, you need to take action to update the neighbors for efficient routing. The neighbors update format is [<node id>, <node id>]. You also need to keep track of the local position of nodes and update them accordingly. The position update is provided as [x,y,z] coordinates. Regarding the network, there are 3 nodes on the mesh network named node1, node2 and node3. The mesh network is set to communicate on channel 36 to start. Based on the network observations that you will receive, you are required to choose the best action from the valid action set to keep up the performance of network and to protect it against security threats. Please, answer that you understood the context.

User prompt (Instructions)

The network observations are: Network Status from Node1 Best Neighbors List is [2, 3].
The valid actions set contains (#):
Disconnect all nodes from node 2 # Disconnect all nodes from node 3 # Switch all nodes to channel 36
Switch all nodes to channel 37 # Switch all nodes to channel 38 # Switch all nodes to channel 39
Switch all nodes to channel 40 # Switch all nodes to channel 41 # Switch all nodes to channel 42
Switch all nodes to channel 43 # Switch all nodes to channel 44 # Switch all nodes to channel 45
Switch all nodes to channel 46 # Update Neighbors of node 1 # Update Neighbors of node 2
Update Neighbors of node 3 # Set Network Throughput to 0.1 Mb/s for all nodes
Set Network Throughput to 2 Mb/s for all nodes # Set Network Throughput to 10 Mb/s for all nodes
Update Local Position of node 1 # Update Local Position of node 2 # Update Local Position of node 3
No Action.
INSTRUCTIONS:
- You MUST choose only ONE action from the valid action set.
- You MUST identify your chosen action by the tag <ACTION>your choosen action</ACTION>.
- Do NOT respond with any other additional text, and you CANNOT decline to take an action.

LLM response

<ACTION>Update Neighbors of node 1</ACTION>

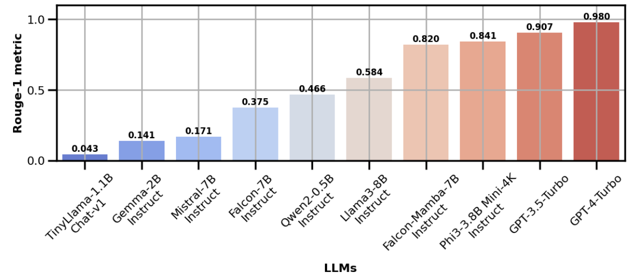
metrics: *ROUGE-1*, *METEOR*, and *BLEU* scores. These metrics were calculated by comparing the LLM-generated responses to preferred labeled responses for each network state. The results are presented in Fig. 2. Our findings show that GPT models (specifically `GPT-3.5` and `GPT-4 Turbo`) consistently produce high-quality responses to the instructed tasks, even with zero-shot prompting. This performance can be attributed to their large-scale pre-training and extensive knowledge base. On the open-source side, most models demonstrated lower performance on the task despite well-formatted *system* and *user* prompts. This is likely due to their smaller scale per-training and size. Interestingly, some recent open-source models, such as `Phi3-3.8B Mini-4K` and `Falcon-Mamba-7B`, achieved notably higher performance comparable to the GPT models. Furthermore, the contrast between their high *ROUGE-1* and *METEOR* scores and their average *BLEU* score suggests that while there is some discrepancy with the exact match to the preferred labeled responses, these models grasp the core task and context. Indeed, as demonstrated in Table 2 for `Falcon-Mamba-7B`, the quality of responses from these open-source LLMs is sensitive to minor changes in prompt format. Therefore, further fine-tuning and alignment are necessary to improve their performance on the mesh network management task.

Table 2: Sensitivity of Falcon-Mamba-7B to prompt format

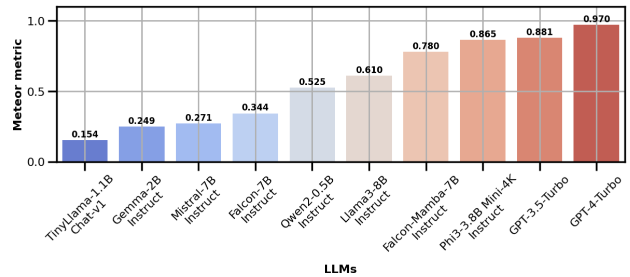
Falcon-Mamba-7B	ROUGE-1	METEOR	BLEU
Prompt ends with '\n'	0.82	0.78	0.58
Prompt ends without '\n'	0.38	0.38	0.54
Prompt ends with '\n\n'	0.67	0.66	0.51

4 CONCLUSION

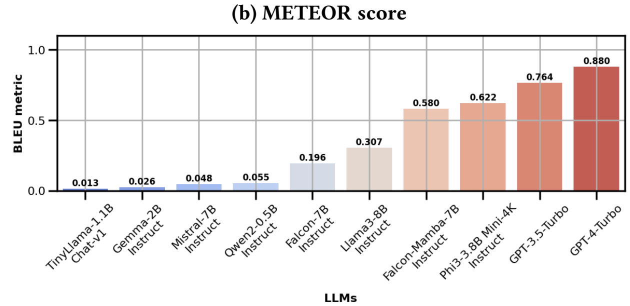
In this paper we have explored the use of LLMs for real-time network management systems. Despite their high performance, the use of GPT models for network management tasks presents significant challenges due to limited access, high usage costs, and privacy concerns. However, our benchmarking results clearly indicate that



(a) ROUGE-1 score



(b) METEOR score



(c) BLEU score

Figure 2: Benchmarking results across standard LLM metrics.

some recent open-source models, such as `Phi3-3.8B Mini-4K` and `Falcon-Mamba-7B`, offer promising perspectives, thanks to their smaller size, easy accessibility, and impressive zero-shot performance. Nevertheless, to meet the strict requirements for reliability and resiliency in wireless network management, these open-source models will require further fine-tuning and alignment to be fully effective for this task.

REFERENCES

[1] Hirenkumar Kamleshbhai Mistry et al. Artificial Intelligence For Networking. *Educational Administration: Theory and Practice*, 2024.
[2] Uchenna Joseph Umoga et al. Exploring the potential of AI-driven optimization in enhancing network performance and efficiency. *Magna Scientia Advanced Research and Reviews*, 2024.
[3] Chenghao Huang et al. Large Foundation Models for Power Systems. *ArXiv*, 2023.
[4] Wentao Zhang et al. A Multimodal Foundation Agent for Financial Trading: Tool-Augmented, Diversified, and Generalist. In *Proc. of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2024.
[5] N. G. Kuftinova et al. Large Language Model in Suburban Transport Data Management. In *Proc. of the Systems of Signals Generating and Processing in the Field of on Board Communications*, 2024.
[6] Hang Zou et al. TelecomGPT: A Framework to Build Telecom-Specfic Large Language Models. *ArXiv*, 2024.
[7] Abdelkader Mekrache et al. LLM-enabled Intent-driven Service Configuration for Next Generation Networks. In *Proc. of the IEEE 10th International Conference on Network Softwarization (NetSoft)*, 2024.
[8] Changjie Wang et al. NetConfEval: Can LLMs Facilitate Network Configuration? In *Proc. of the ACM on Networking*, 2024.