Mission Explainable: From Feature Attribution to Mitigation in 5G Anomaly Detection

Federica Uccello*, Imtiaz Karim‡, Kazi Samin Mubasshir†, Elisa Bertino†*, Simin Nadjm-Tehrani*
*Department of Computer and Information Science, Linköping University, Sweden
†Department of Computer Science, Purdue University, USA
‡Department of Computer Science, The University of Texas at Dallas, USA
Emails: federica.uccello@liu.se; imtiaz.karim@utdallas.edu;
{kmubassh, bertino}@purdue.edu; simin.nadjm-tehrani@liu.se

***Abstract*—Next-generation networks' complexity translates to a broad attack surface, increasingly hard to monitor and protect. In the ongoing cybersecurity arms race, as attackers exploit Artificial Intelligence (AI) to design new threats, defenders must operate at the same level. While AI-based anomaly detection has shown great promise, its interpretability is often hindered by the black box nature of modern models. Explainable AI (XAI) methods, starting from feature attribution, can address this challenge by providing insights into the model's decision-making process. Yet, it remains unclear to what extent XAI can help analysts interpret alerts and guide mitigation actions. In this paper, we explore the use of large language models (LLMs) as an additional interpretive layer in the anomaly detection pipeline. We propose ROXAS (Reasoning Over eXplained AnomalieS), a methodology that combines anomaly detection via an XGBoost regressor trained solely on benign data, logic-based feature attribution for correct interpretation of alerts, and LLM-based guidance to move toward actionable mitigation assistance. The LLM outputs are evaluated against the expert-curated MITRE FiGHT database, showing alignment with best practices in 5G network defense.**

***Index Terms*—5G Networks, Anomaly Detection, Explainable Artificial Intelligence, Large Language Models**

I. Introduction

As 5G networks are increasingly deployed in critical applications, their growing complexity introduces new challenges for ensuring robust and adaptive cybersecurity [1]. Artificial Intelligence (AI)-based anomaly detection has emerged as a promising strategy to tackle these challenges [2]–[4]. However, such models are often perceived as black boxes that lack insight into their decision-making processes.

Explainable AI (XAI) seeks to bridge the interpretability gap as well as other perceived deficiencies, with several approaches explored in the literature [5], [6]. Among interpretability approaches, *feature attribution* methods identify which input features contributed the most to a given output. While feature attribution can reveal *what* features are responsible for an anomaly, it does not tell *how* these insights should inform mitigation actions.

Recent studies highlight the potential of Large Language Models (LLMs) to enhance cybersecurity "reasoning" by deriving human-like insights from complex, heterogeneous data [7]. Building on this promise, we aim to understand whether LLMs can translate low-level insights into operational mitigation guidance. In this work, we explore the integration of feature attribution and LLMs in AI-driven anomaly detection, addressing the following research questions (RQs):

- RQ1: Can an LLM translate feature attribution into operational mitigation guidance?
- RQ2: To what extent do LLM-derived mitigation strategies align with established best practices in 5G network defense?

To address these questions, we propose ROXAS (Reasoning Over eXplained AnomalieS), a methodology to enhance the interpretability and operational relevance of AI-based network defense systems, and evaluate it against multi-step attacks in 5G cellular network.

The remainder of the paper is organized as follows: Section II reviews background concepts; Section III presents the proposed methodology; Section IV introduces the case study; Section V describes implementation details and evaluation strategy; Section VI reports and discusses results; Section VII summarizes related work; finally, Section VIII concludes the paper.

II. Background

A. Anomaly Detection via Regression

Anomaly detectors typically use unsupervised learning on benign data to establish a model of normality [8]–[11]. We employ an XGBoost regressor [12], a highly scalable gradient-boosted decision tree algorithm suitable for network intrusion detection [13]. Although inherently supervised, XGBoost can be adapted for one-class anomaly detection by training exclusively on benign data, as done in the present research. Large prediction errors are treated

979-8-3315-7111-5/26/$31.00 ©2026 IEEE

as anomalies, enabling binary classification of normal and abnormal instances.

B. Logic-based Feature Attribution

XAI provides insight into AI-model decisions through global or local explanations, using statistical or logical methods [14]–[17]. We focus on feature attribution, well-suited for applications where understanding the influence of specific inputs is crucial, such as network security [15]. Specifically, we use VoTE-XAI [18], a logic-based method for tree ensembles that has been shown to have superior performance [19]. It yields minimal, logically sound explanations by identifying feature–value pairs necessary to justify a given prediction.

Formally, an explanation E is *valid* for a prediction $f(c_1,\ldots,c_n) \mapsto d$ if:

$$\bigwedge_{(x_i,c_i)\in E} (x_i = c_i) \implies f(x_1,\ldots,x_n) = d \qquad (1)$$

Where $x_1,\ldots,x_n$ are the features, $c_1,\ldots,c_n$ are their respective values, and d is the model's prediction. An explanation E is *minimal* if removing any element from it invalidates this justification:

$$\forall A \subset E,\ \bigwedge_{(x_i,c_i)\in A} (x_i = c_i) \not\Longrightarrow f(x_1,\ldots,x_n) = d \qquad (2)$$

Where $x_1,\ldots,x_n$ are the features, $c_1,\ldots,c_n$ are their respective values, and d is the model's prediction.

C. Large Language Models

LLMs are transformer-based neural networks trained on massive text corpora to learn probabilistic language representations. Leveraging self-attention mechanisms [20], they capture contextual dependencies and generalize across domains, enabling few-shot and zero-shot learning [21]. Their versatility makes them well-suited as interpretive layers in complex pipelines requiring human-readable analysis and summarization. We use Open-Mistral-7B[1], an open-source LLM accessible via API, due to its efficiency and balanced trade-off between performance and computational cost [7].

D. MITRE FiGHT Database

MITRE Five-G Hierarchy of Threats (FiGHT) [22] is a publicly available knowledge base for 5G network defense that enumerates adversarial techniques (FiGHT Technique IDs, FGT) and corresponding mitigations (FiGHT Mitigation IDs, FGM). Each entry includes detailed technical descriptions covering radio, core, and management planes. In this work, the framework serves as a reference to evaluate whether the LLM-generated mitigation strategies align with recognized best practices.

[1]https://docs.mistral.ai/

III. Proposed Methodology

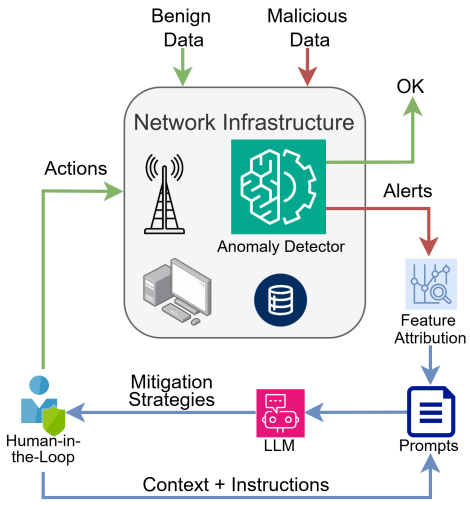

Fig. 1. High-level representation of ROXAS.

ROXAS is structured in four steps (Figure 1):

1) **Anomaly Detection:** We employ an XGBoost regressor trained exclusively on benign data to learn a model of normality. During inference, instances with large prediction error are flagged as alerts.
2) **Feature Attribution**: We apply VoTE-XAI to the alerts from Step 1, computing one minimal explanation per alert, consisting of feature–value pairs that are necessary to justify the model's output.
3) **Prompt Construction**: We use the output from Step 2 to construct structured prompts for the LLM. Each prompt includes a contextual description of the system's role (network analyst), a brief instruction specifying the reasoning task, and the list of anomalous features and their values.
4) **Mitigation Actions**: Finally, we prompt the Open-Mistral-7B model to translate low-level feature attributions into high-level, human-readable guidance, using the output of Step 3.

IV. Case Study: Fake Base Station in 5G Networks

Our dataset captures Layer-3 network traces from diverse attacker capabilities and mobility settings, enabling realistic modeling of Fake Base Stations and multi-step attacks. The traces include Non-Access Stratum (NAS) and Radio Resource Control (RRC) protocol messages. We focus on NAS-level features only - see [23] for details.

A. Data Pre-processing

To improve detection performance, multiple data pre-processing steps were performed. First, low-variance features carrying little informational value were removed. To address redundancy, pairs of highly correlated features were compared, and the feature with the lower mutual information score was discarded. This process reduced the feature space from 478 to 117 features.

As the model does not perform classification, strict class balance is unnecessary. To enhance generalization, 3,000 additional benign samples were synthetically added by sampling existing benign instances and adding small Gaussian perturbations to their numerical features. The noise was scaled by each feature's standard deviation to preserve realistic value ranges, while categorical and binary features were left unchanged. The number of samples was determined empirically based on validation performance.

Finally, all features were standardized using z-score normalization to ensure comparable numerical ranges and to prevent high-magnitude features from disproportionately influencing the model's predictions.

B. Attack Selection

The dataset includes 21 attacks across four macro-categories: Denial of Service (DoS), Bidding Down (a category of selective DoS), Location Tracking, and Battery Drain. We selected one representative class per macro-category, based on sufficient representation in the dataset and diversity, measured via pairwise cosine distances between class centroids. The final selection included RRC replay attack for DoS, Bidding Down with Service Reject, Location Tracking via Measurements Report, and Energy Depletion Attack.

C. Evaluation Goals

For Step 1 and 2 of ROXAS (Section III), the anomaly detector is applied to the test data to produce alerts that serve as input for subsequent analysis. Its performance reflects how accurately these alerts correspond to genuine anomalies. Then, VoTE-XAI produces explanations that are provably correct with respect to the model.

Finally, for Steps 3 and 4, the aim is to assess whether the LLM-generated outputs are operationally meaningful and aligned with known countermeasures for the corresponding anomalies.

V. Implementation and Evaluation Strategy

All the steps were implemented in Python, and executed in a Debian-based virtual machine equipped with an Intel Core Ultra 7 155U CPU and 32 GB RAM. Source code and data are available on GitHub[2].

Steps 1 and 2 of ROXAS were applied in accordance with existing methodology [19]. The hyperparameter tuning and performance of the detector, both of which form the starting point of this work, are summarized in Table I. The metrics definition is standard[3]. The model achieved an overall precision of 0.96, recall of 0.97, and F1-score of 0.97, correctly identifying the majority of malicious instances with a low false positive rate.

TABLE I
Anomaly Detector Configuration and Performance.

Hyperparameters	
Number of estimators	200
Max depth	10
Learning rate	0.1
Performance	
TP	23,533
FP	955
TN	2,950
FN	722
Precision	0.96
Recall	0.97
F1-Score	0.97

A. Prompt Construction

In Step 3, the prompt template was refined iteratively during initial testing. A context file was defined as follows:

- Context: A known fake base station has been detected in the 5G network.
- An anomaly detection system has flagged a suspicious NAS (Non-Access Stratum) trace that may be related to this activity.
- Input: You are provided with the most anomalous feature-value pairs for one suspicious trace.
- Task: Provide actionable mitigation recommendations to restore normality.

Then, three representative true positive alerts were selected for each macro-category, and their feature attributions were used to construct prompts through an automated script.

Each message sent to the LLM consisted of: (i) a *system* message defining the analyst's role, and (ii) a *user* message including the above context, the list of anomalous feature–value pairs, and explicit instructions to summarize the likely cause and propose a mitigation checklist limited to 5G NAS/RRC and core/RAN configuration. The model was also instructed to specify timers, counters, rules, or KPIs where applicable, and to state assumptions when evidence was insufficient.

[2]https://github.com/FedeU95/ROXAS

[3]https://scikit-learn.org/stable/modules/model_evaluation.html

B. Mitigation Actions

Step 4 was implemented using Open-Mistral-7B. The LLM was queried via its API with the context file and a prompt for each anomaly and attribution from Steps 2 and 3.

The generated outputs were first reviewed for internal consistency within each macro-category as a sanity check. Subsequently, each LLM response was evaluated via mapping onto a mitigation from the expert-curated MITRE FiGHT framework.

Each selected attack was mapped with the corresponding FGT. Then, each LLM-suggested action was manually examined and associated with the most relevant FGM entry by comparing its technical description against the FiGHT mitigation taxonomy.

VI. Results and Discussion

Table II shows the mapping between the LLM-suggested actions and MITRE FiGHT mitigation, where applicable. A minority of LLM-generated actions had no direct FGM counterpart, but aligned with broader defensive principles (e.g. recommendations to collaborate with other network operators, implement user education programs for awareness).

For Bidding Down, the LLM correctly identified the anomaly as an attempt to manipulate NAS procedures to cause service rejection or DoS. Its proposals to verify message integrity, adjust parameters, and restrict reconnections mapped to FGM5002 (Discard RAN signaling received without integrity protection) and FGM5006 (Restrictive user profile), both countermeasures to the FGT1562.501 technique (Impair Defenses: Bid-Down UE).

For Energy Depletion, the generated actions (blocking or whitelisting suspicious NAS-IDs, enforcing anomaly detection rules, and monitoring attach/detach activity) mapped to FGM1010 (Deploy compromised device detection method) and FGM1040 (Behavior prevention on endpoint), both mitigating FGT1203.502 (Exploitation for Client Execution: Baseband API).

Location Tracking results were highly consistent, linking the anomalies to user impersonation or information extraction. The LLM repeatedly suggested increasing RRC re-establishment timers, enforcing frequent authentication and key-agreement procedures, and strengthening encryption. These correspond to FGM5006 (Restrictive user profile) and FGM1041 (Encrypt sensitive information), which defend against FGT5012.002 (Locate UE: UE Measurement Reports).

For RRC Replay, recommendations such as isolating the affected UE, limiting repeated connection attempts, and monitoring signaling rates all mapped to FGM5509 (Screen signaling and user-plane messages), associated with FGT1498.501 (Network DoS: Flooding Core Network Component).

A. Insights from Evaluation

Results show that our anomaly detection model, which was only trained on benign data, achieves a binary classification performance comparable to prior work [23]. This is interesting since adopting an unsupervised approach is a more pragmatic approach in real applications and reduces the need for labeled attack class data.

These outcomes indicate that we can start the explanation and mitigation approach with a a lightweight, regression-based model that effectively captures abnormal network dynamics and generalizes to unseen attack instances.

The LLM results are indicative of the novel contribution of this work. The outcomes of steps 3 and 4 demonstrate that structured prompts derived from feature attribution can yield technically coherent and operationally meaningful mitigation guidance. This supports the premise that LLMs, guided by feature attribution, can serve as an intermediate interpretative layer in the detection pipeline (RQ1). Note that the LLM's guidance is largely driven by the quality of the explanatory input and, implicitly, by the model used for detection. Finally, most LLM-generated suggestions could be mapped with MITRE FiGHT mitigation, suggesting that LLM-assisted guidance can translate feature attributions into high-level, context-aware hints so the operator, aligned with established best practices when placed in the context of a 5G network defense (RQ2).

B. Limitations and Future Work

Despite the promising results, our approach has limitations. First, the evaluation focused primarily on user-level attacks rather than infrastructure-level threats, not focusing on the generalizability of the findings to broader scenarios. Second, the LLM we used is a general-purpose, mid-sized model not specifically trained on 5G or network security data. Third, the qualitative assessment of the LLM outputs was not confirmed through human-in-the-loop validation by domain experts, which would strengthen the evidence of the operational soundness of the generated suggestions.

Future work will address these limitations by extending ROXAS to infrastructure-level attack datasets currently under collection, experimenting with larger and fine-tuned LLMs specialized for 5G network contexts, and integrating expert-driven evaluation loops to further assess the accuracy and applicability of LLM-generated recommendations.

TABLE II
MAPPING BETWEEN LLM-SUGGESTED MITIGATIONS AND MITRE FIGHT DEFENSIVE MEASURES.

Attack Class	**LLM-Suggested Mitigations**	**Mapped MITRE FiGHT Mitigation IDs**	**MITRE FiGHT Technique ID**
Bidding Down with ServiceReject	1. Analyze NAS and RRC messages for integrity; 2. Tighten NAS EPS EMM parameters; 3. Configure NAS timers to limit exposure; 4. Blacklist suspicious IMSIs to prevent reconnection; 5. Deploy IDS/IPS for signaling integrity.	FGM5002 (1,5); FGM5006 (2,3,4)	FGT1562.501
Energy Depletion	1. Block suspicious NAS-ID or TMSI to prevent unauthorized access; 2. Update NAS-ID whitelist to restrict allowed devices; 3. Implement NAS/RRC anomaly detection rules; 4. Monitor KPIs (NAS-ID distribution attach/detach rates); 5. Update core-network security policies and AAA rules.	FGM1010 (1, 2, 3,4); FGM1040 (5)	FGT1203.502
Location Tracking	1. Increase RRC re-establishment timer to slow reconnection; 2. Enforce frequent authentication and key agreement (AKA); 3.Apply IDS rules to detect abnormal NAS/RRC identifiers; 4. Enforce strong encryption and authentication at RAN and core.	FGM5006 (1, 3); FGM1041 (2, 4)	FGT5012.002
RRC Replay	1. Isolate affected UE from network; 2. Increase RRC re-establishment timer to limit repeated requests; 3. Implement IDS rules to detect repeated or out-of-sequence NAS/RRC messages; 4. Monitor NAS message rates and anomalies; 5. Apply configuration updates and firmware patches to prevent signaling abuse.	FGM5509 (1-5)	FGT1498.501

VII. RELATED WORK

Research leveraging LLMs for network security is still in early stages, with many studies presenting preliminary experiments. Mandal et al. [24] propose an LLM-based approach for identifying vulnerable code segments and suggesting mitigation strategies. Wang et al. [25] introduce an LLM-based framework for DDoS mitigation, that leverages complete contextual information about the attack scenario to generate device-specific configuration commands.

Dong et al. [26] explore the potential of LLMs to automate the refinement of cellular network specifications, while Dayaratne et al. [27] propose an O-RAN–compliant, LLM-based framework for DoS detection in 5G and beyond.

Other approaches have investigated federated and edge-based LLM applications. Rezaei et al. [28] propose a federated framework combining LLMs and local learning agents for adaptive anomaly detection and robustness to adversarial AI threats. Bani-Melhem et al. [29] propose an explainable LLM-based APT detection system for 6G security, where explainability refers to translating raw network data into interpretable natural language before classification.

In contrast to previous works, our approach integrates logic-based feature attribution with LLM-driven guidance for automated interpretation and mitigation of network anomalies. Moreover, we evaluate the generated recommendations against the expert-curated MITRE FiGHT framework to provide a standardized measure of alignment with established 5G defense practices.

VIII. CONCLUSION

This work introduced ROXAS, a methodology that combines anomaly detection and feature attribution

with LLM to improve interpretability and operational relevance in 5G network defense. Using an XGBoost regressor, logic-based feature attribution, and structured LLM prompting, we showed that low-level explanations can be translated into coherent and actionable mitigation guidance, aligned with standard practices.

ACKNOWLEDGEMENT

This work was carried out within the NEST project *AIR*^{2}, which is partially supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation. It was also supported by ELLIIT, the Excellence Center at Linköping - Lund in Information Technology, and the US NSF under grants 2112471 and 2229876.

REFERENCES

[1] P. Benlloch-Caballero, Q. Wang, and J. M. A. Calero, "Distributed dual-layer autonomous closed loops for self-protection of 5G/6G IoT networks from distributed denial of service attacks," *Computer Networks*, vol. 222, p. 109526, 2023.

[2] M. E. Morocho-Cayamcela, H. Lee, and W. Lim, "Machine learning for 5G/B5G mobile and wireless communications: Potential, limitations, and future directions," *IEEE access*, vol. 7, pp. 137 184–137 206, 2019.

[3] L. A. Garrido, A. Dalgkitsis, G. Famitafreshi, A. Siokis, K. Ramantas, and C. Verikoukis, "An experimental platform of a beyond-5Gnetwork with machine learning integration," in *GLOBECOM 2023-2023 IEEE Global Communications Conference*. IEEE, 2023, pp. 62–67.

[4] U. K. Lilhore, S. Dalal, and S. Simaiya, "A cognitive security framework for detecting intrusions in IoT and 5G utilizing deep learning," *Computers & Security*, vol. 136, p. 103560, 2024.

[5] V. Hassija, V. Chamola, A. Mahapatra, A. Singal, D. Goel, K. Huang, S. Scardapane, I. Spinelli, M. Mahmud, and A. Hussain, "Interpreting black-box models: a review on explainable artificial intelligence," *Cognitive Computation*, vol. 16, no. 1, pp. 45–74, 2024.

[6] W. Saeed and C. Omlin, "Explainable AI (XAI): A systematic meta-survey of current challenges and future opportunities," *Knowledge-Based Systems*, vol. 263, p. 110273, 2023.

[7] M. A. Ferrag, F. Alwahedi, A. Battah, B. Cherif, A. Mechri, N. Tihanyi, T. Bisztray, and M. Debbah, "Generative ai in cybersecurity: A comprehensive review of llm applications and vulnerabilities," *Internet of Things and Cyber-Physical Systems*, 2025.

[8] D. Samariya and A. Thakkar, "A comprehensive survey of anomaly detection algorithms," *Annals of Data Science*, vol. 10, no. 3, pp. 829–850, 2023.

[9] A. Pinto, L.-C. Herrera, Y. Donoso, and J. A. Gutierrez, "Enhancing critical infrastructure security: Unsupervised learning approaches for anomaly detection," *International Journal of Computational Intelligence Systems*, vol. 17, no. 1, p. 236, 2024.

[10] Z. Zhang, Z. Zhao, X. Zhang, and X. Chen, "DA2: Distribution-agnostic adaptive feature adaptation for one-class classification," *Computer Vision and Image Understanding*, vol. 251, p. 104256, 2025.

[11] J. Ye, Z. Tan, Y. Hu, X. Yang, G. Cheng, and K. Huang, "Disentangling tabular data towards better one-class anomaly detection," in *Proceedings of the Thirty-Ninth AAAI Conference on Artificial Intelligence and Thirty-Seventh Conference on Innovative Applications of Artificial Intelligence and Fifteenth Symposium on Educational Advances in Artificial Intelligence*. AAAI Press, 2025.

[12] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.

[13] E. Edozie, A. N. Shuaibu, B. O. Sadiq, and U. K. John, "Artificial intelligence advances in anomaly detection for telecom networks," *Artificial Intelligence Review*, vol. 58, no. 4, p. 100, 2025.

[14] H. Sun, Y. Liu, A. Al-Tahmeesschi, A. Nag, M. Soleimanpour, B. Canberk, H. Arslan, and H. Ahmadi, "Advancing 6G: Survey for explainable AI on communications and network slicing," *IEEE Open Journal of the Communications Society*, vol. 6, pp. 1372–1412, 2025.

[15] H. Ouifak and A. Idri, "A comprehensive review of fuzzy logic based interpretability and explainability of machine learning techniques across domains," *Neurocomputing*, p. 130602, 2025.

[16] G. Schwalbe and B. Finzel, "A comprehensive taxonomy for explainable artificial intelligence: a systematic survey of surveys on methods and concepts," *Data Mining and Knowledge Discovery*, vol. 38, no. 5, pp. 3043–3101, 2024.

[17] M. Mersha, K. Lam, J. Wood, A. K. Alshami, and J. Kalita, "Explainable artificial intelligence: A survey of needs, techniques, applications, and future direction," *Neurocomputing*, vol. 599, p. 128111, 2024.

[18] J. Törnblom, E. Karlsson, and S. Nadjm-Tehrani, "Finding minimum-cost explanations for predictions made by tree ensembles," *In Print*, 2025. [Online]. Available: diva2:1953715

[19] F. Uccello and S. Nadjm-Tehrani, "Investigating feature attribution for 5G network intrusion detection," 2025. [Online]. Available: https://arxiv.org/abs/2509.10206

[20] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[21] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell *et al.*, "Language models are few-shot learners," *Advances in neural information processing systems*, vol. 33, pp. 1877–1901, 2020.

[22] "MITRE FIGHT," https://fight.mitre.org/, 2025, online.

[23] K. S. Mubasshir, I. Karim, and E. Bertino, "Gotta detect 'em all: Fake base station and multi-step attack detection in cellular networks," in *Proceedings of the 34th USENIX Security Symposium*, 2025.

[24] U. Mandal, S. Shukla, A. Rastogi, S. Bhattacharya, and D. Mukhopadhyay, "μlam: A LLM-powered assistant for real-time micro-architectural attack detection and mitigation," in *Proceedings of the 43rd IEEE/ACM International Conference on Computer-Aided Design*, 2024, pp. 1–9.

[25] T. Wang, X. Xie, L. Zhang, C. Wang, L. Zhang, and Y. Cui, "ShieldGPT: An LLM-based framework for DDoS mitigation," in *Proceedings of the 8th asia-pacific workshop on networking*, 2024, pp. 108–114.

[26] J. Dong, T. Zhang, F. Yan, Y. Li, H. Li, and H. Qiu, "Can large language models automate the refinement of cellular network specifications?" *arXiv preprint arXiv:2507.04214*, 2025.

[27] T. Dayaratne, N. D. Pham, V. Vo, S. Lai, S. Abuadbba, H. Suzuki, X. Yuan, and C. Rudolph, "From description to detection: LLM based extendable O-RAN compliant blind dos detection in 5G and beyond," *arXiv preprint arXiv:2510.06530*, 2025.

[28] H. Rezaei, R. Taheri, and M. Shojafar, "FedLLMGuard: A federated large language model for anomaly detection in 5G networks," *Computer Networks*, p. 111473, 2025.

[29] S. Bani Melhem, M. Golec, A. Alwarafy, and Y. Khamayseh, "LENS: Lightweight and explainable LLM-Based APT detection at the edge for 6G security," *IEEE Access*, vol. 13, pp. 172 402–172 415, 2025.