LLM-Driven APT Detection for 6G Wireless Networks: A Systematic Review and Taxonomy

Muhammed Golec[1,2], Yaser Khamayseh[3], Suhib Bani Melhem[4], Abdulmalik Alwarafy[5]

[1]School of Electronic Engineering and Computer Science, Queen Mary University of London, UK
[2]Computer Science Department, Abdullah Gul University, Kayseri, Turkey
[3]The College of Technological Innovation, Zayed University, UAE
[4] College of Engineering, Al Ain University, Abu Dhabi, UAE
[5] College of Information Technology, UAE University, Al Ain, UAE

Preprint submitted to Computer Science Review (Elsevier), May 2025

Abstract

Sixth Generation (6G) wireless networks, which are expected to be deployed in the 2030s, have already created great excitement in academia and the private sector with their extremely high communication speed and low latency rates. However, despite the ultra-low latency, high throughput, and AI-assisted orchestration capabilities they promise, they are vulnerable to stealthy and long-term Advanced Persistent Threats (APTs). Large Language Models (LLMs) stand out as an ideal candidate to fill this gap with their high success in semantic reasoning and threat intelligence. In this paper, we present a comprehensive systematic review and taxonomy study for LLM-assisted APT detection in 6G networks. We address five research questions, namely, semantic merging of fragmented logs, encrypted traffic analysis, edge distribution constraints, dataset/modeling techniques, and reproducibility trends, by leveraging most recent studies on the intersection of LLMs, APTs, and 6G wireless networks. We identify open challenges such as explainability gaps, data scarcity, edge hardware limitations, and the need for real-time slicing-aware adaptation by presenting various taxonomies such as granularity, deployment models, and kill chain stages. We then conclude the paper by providing several research gaps in 6G infrastructures for future researchers. To the best of our knowledge, this paper is the first comprehensive systematic review and classification study on LLM-based APT detection in 6G networks.

1 Introduction

The rapid development of wireless technologies increases expectations for 6G networks with ultra-low latency and artificial intelligence-based orchestration architecture [1]. To meet these expectations, 6G networks operate with a heterogeneous architecture where many layers, such as physical and network layers, work together, which means a larger attack surface [2]. The wide attack surface in 6G systems requires that measures be taken against Advanced Persistent Threats (APTs), one of the hidden and long-stage attack methods that are difficult to detect with traditional detection mechanisms [3].

Large Language Models (LLMs) with semantic and contextual reasoning features are one of the most promising developments that can be used against APTs [4]. In particular, it can be used for the detection of APT in 6G networks by analyzing fragmented logs and increasing situational awareness [5]. Despite this potential, there is no comprehensive taxonomy or systematic analysis in the literature on LLM-based APT detection for 6G networks.

To the best of our knowledge, this paper is the first comprehensive systematic review and classification study on LLM-based APT detection in 6G networks. As a result of the current studies examined by the authors using identification, screening, eligibility, and inclusion (snowballing) techniques, 142 articles were analyzed. Our aimed is to synthesize the intersection of LLM architectures, APT lifecycle modeling, and 6G-specific security challenges and provide insights for future research.

1.1 Motivation and Contributions

LLMs and 6G technologies are very recent research areas and their intersection in cyber threat detection, such as APT attacks, has not been sufficiently investigated in the literature. Existing studies are scattered across either LLM-based cybersecurity or 6G network security issues. Furthermore, 6G networks are still in their infancy (expected to become widespread after the 2030s) and contain obstacles for AI and rule-based systems due to issues such as a fragmented structure of source data and end device limitations [6]. For all these reasons, a detailed investigation should be conducted to explore the potential of LLMs in providing explainable detection mechanisms throughout the 6G infrastructure. The main contributions of this paper can be summarized as follows:

- We present the first Systematic Literature Review (SLR)-based review focusing on LLM-enabled APT detection in 6G networks. To do this, we searched more than 300 most recent and relevant papers in academic and industrial databases between 2018-2025. As a result of the systematic analysis (Kitchenham's SLR approach and Petersen's Systematic Mapping Study (SMS) [7, 8]), the most relevant 142 papers in the field were obtained.
- We define five-point research questions to conduct the systematic review (Section 4.2). In line with these questions: (i) Semantic correlation of fragmented logs generated in 6G networks and how it can be used for LLMs threat detection (Section 5.1), (ii) Limitations of 6G encrypted channels and how it can address LLMs visibility and reasoning challenges (Section 5.2), (iii) Challenges of deploying LLM to edge nodes on 6G networks and optimization techniques for these challenges (Section 5.3), (iv) Datasets and modeling techniques used in LLM-based APT detection studies (Section 5.4), and (v) Exploration of publication trends, platform distribution, and reproducibility for LLM-focused APT research (Section 5.5).
- LLM deployment models, threat lifecycle stages, optimization strategies for edge inference, and taxonomy studies for dataset types are presented.
- Research gaps, such as explainability gaps, dataset scarcity, and 6G orchestration risks, are highlighted through critical analysis. And future directions, such as slice-aware XAI pipelines and unified demand tuning techniques, are highlighted.
- A comparison of this paper with 16 previous reviews in the literature. The comparison is made to highlight the novelty and necessity of the paper.

1.2 Article Organization

Figure 1 shows the organizational chart for this paper. Section 2 provides a comparison with related surveys to highlight the novelty of the paper. Section 3 explains the basic background of APTs, 6G networks, and LLMs, and explains their roles in cybersecurity. Section 4 explains the methodological structure, such as the article selection methods and research questions for this systematic review. Section 5 provides an in-depth analysis of five key research questions. Section 6 indicates open challenges and future directions for researchers in the relevant research area.

Section 7 concludes the paper by summarizing the findings and emphasizing the importance of LLM-based APT detection in 6G.

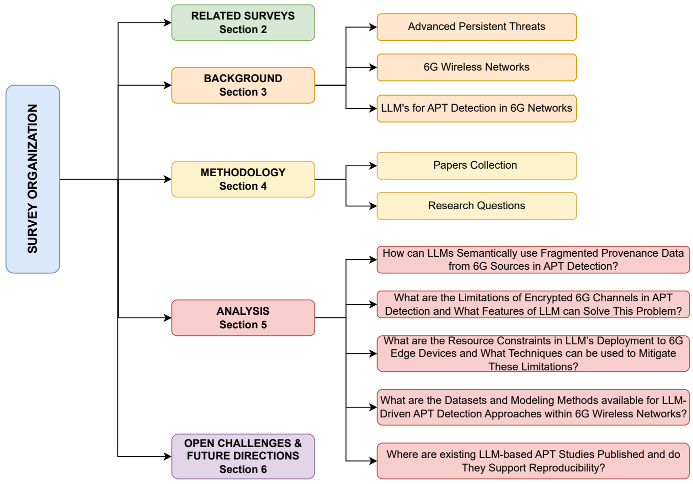

Figure 1: The Organization of the Survey

2 RELATED SURVEYS

The use of LLMs in the detection and prevention of APTs, which are potential cybersecurity threats in 6G, is still an area that needs to be investigated. The main reason for this is that there are limited datasets in the literature on APTs and 6Gs can only be modeled simulation-based. When the literature is examined, it is seen that although there are surveys focused on 6G, APT, LLMs, and LLM-based security, however, there is no systematic review and taxonomy study that addresses LLM, APT, and 6G in a combined manner.

LLM-Focused Cybersecurity Surveys: Several recent studies in the literature have examined the role of LLMs in the field of cybersecurity. Hassanin et al. [9] provide an overview of the role of LLMs in applications such as threat intelligence and phishing detection in their review. In [10, 11, 12, 13], they examine the architectures used for LLM-based attack detection and threat analytics in a more systematic way. Zuo et al. [14] presented an analysis study examining LLM's usage for APTs. This study investigated the semantic augmentation of LLM's (such as GPT-4o) origin logs for APT detection. However, this study is superficial, not a survey or taxonomy, and does not include the 6G context. In another study, the authors present a review of language models (including APT), but do not include any information about 6G and do not use a formal methodology such as SLR [15]. Although LLM's sheds light on the applications in cybersecurity, none of these studies cover 6G and its limitations and opportunities.

APT-Focused Surveys: Some of the literature studies investigated APT detection using DL and rule-based learning. In [16, 17], classifications and threat lifecycle analyses of APTs are examined, while in [18] DL-based cyber attack detection systems (partially addressing APT) are investigated. Although all these survey studies partially or in detail mention APTs, none of them provide detailed information about LLM-based approaches or 6Gs (such as network layer dynamics).

6G Focused Surveys: Another area of survey research in the literature examines the technical foundations of 6G. Shen et al. [19] covers five main aspects of 6G (such as spectrum and positioning) in detail. In [20, 21, 22, 23], important components in 6G (such as IoT integration and federated learning) are comprehensively examined. In another survey study, Sun et al. [24] investigate the importance and use of explainable AI (XAI) in 6G network slicing and vehicle contexts. However, none of these studies address LLM and security issues.

Among the reviewed literature studies, [10, 11, 12, 13, 18] examine model architectures and use cases in detail in their focused research topic using systematic research methodology such as PRISMA. Furthermore, some of the studies [11, 13] provide binary taxonomies of cybersecurity tasks, while in [20, 19, 21, 23] they strongly address 6G at the architecture and protocol level. However, none of the reviewed articles address LLM, APT, and 6G in a unified manner.

2.1 Critical Analysis

Table 1 provides a comparison of 16 recent survey studies with this paper. When the table is examined, it is seen that this paper fills the following three critical gaps:

- **Combining consideration of LLM, APT, and 6G**: None of the reviewed studies simultaneously address the intersection of LLM-based threat detection, APT lifecycle modeling, and 6G network features.
- **Providing a detailed taxonomy for APT detection**: The vast majority of studies are in the form of a general survey, and those that do include a taxonomy lack details such as the lifecycle of APTs.
- **Providing a comparative synthesis across fields**: Few of the reviewed studies include multidimensional comparisons (such as methodology, model types). The lack of such comparative syntheses makes it difficult to assess the overlap and gaps between the topics covered by the survey.

Table 1: Comparison of Our Systematic Review and Taxonomy with Existing Survey Studies

Paper	Focus Area	6G-Specific	APT-Specific	LLM-Specific	Type	SLR Methodology	Publisher	Year
[9]	General Cyber Defence	✗	✗	✓	Review	✗	Arxiv	2024
[10]	LLMs in Cybersecurity	✗	✓	✓	Systematic Review	✓	Arxiv	2024
[11]	IDS with Transformers & LLMs	✗	✓	✓	Review and Taxonomy	✓	Elsevier	2024
[12]	LLMs in Cybersecurity	✓	✓	✓	Systematic Survey	✓	IEEE	2025
[13]	LLMs in Cyber Threat Detection	✗	✓	✓	Systematic Review	✓	Elsevier	2024
[14]	LLM-Augmented Provenance for APT Detection	✗	✓	✓	Research Paper	✗	Sandia TR	2025
[15]	PLMs/LLMs in Cybersecurity	✗	✓	✓	Review	✗	IEEE (ISDFS)	2024
[16]	APT Analysis & Countermeasures	✗	✓	✗	Review and Taxonomy	✗	Springer	2019
[17]	APT Detection Techniques	✗	✓	✗	Survey	✓	TechScience (CMC)	2024
[18]	DL Techniques for IDS	✗	✓	✗	Systematic Survey	✓	ACM	2025
[19]	6G Architectures & Networking	✓	✗	✗	Survey	✗	ACM CSUR	2023
[20]	Federated Learning in 5G/6G Cybersecurity	✓	✗	✗	Comprehensive Survey	✓	IEEE	2025
[21]	6G and IoT Integration	✓	✗	✗	Comprehensive Survey	✗	IEEE	2022
[22]	Optimization & Performance of LIS in 6G	✓	✗	✗	Survey	✗	IEEE Access	2020
[23]	6G Technologies & Architectures	✓	✗	✗	Survey	✗	IEEE	2022
[24]	Explainable AI for 6G	✓	✗	✗	Systematic Survey	✓	IEEE OJ-COMS	2025
Our Paper	**LLMs for APT Detection in 6G**	✓	✓	✓	**Systematic Review and Taxonomy**	✓	**Computer Science Review (Planned)**	**2025**

3 BACKGROUND

This section provides the basic background necessary for the reader to better understand the concepts related to LLM-based APT detection in 6G networks.

3.1 Advanced Persistent Threats

Advanced Persistent Threats (APT) are one of the most effective cyber attacks known due to their characteristics, such as stealth and longevity. This subsection defines APT and explains its key characteristics, lifecycle, and attacker behaviors (TTPs). Then, a comparison of traditional attacks and APTs is provided for APT.

3.1.1 Key Characteristics of APT

APTs use multiple vectors to gain long-term access to an IT environment and are like an attacker with significant expertise and resources [25]. They have three basic characteristics [26]:

- **Advanced**: These attacks specialize in zero-day attacks and tactics to evade detection.
- **Persistent**: They encourage new APT attacks by leaving backdoors in the systems they penetrate.
- **Threat**: They carry out attacks such as espionage, sabotage, or exfiltration of critical data from the systems.

3.1.2 The Lifecycle of APTs

The lifecycle for APT attacks is shown in figure 2 and can be summarized in five basic stages [16, 17, 15]:

- **Reconnaissance**: This is the first stage of the attack, and information about the target is collected (Open Source Scanning (OSINT)), and system vulnerabilities and weaknesses are investigated.
- **Initial Intrusion**: The entry into the target system is achieved through methods such as phishing and malware.
- **Command and Control (C2)**: Preparation of the infrastructure to communicate with the APT inserted into the target system (such as backdoor channels).
- **Lateral Movement**: Infiltration of other devices connected to the same network within the system and detection of high-value targets.
- **Data Exfiltration**: The final stage involves malicious operations such as exfiltration of data in the target system using APT and system sabotage.

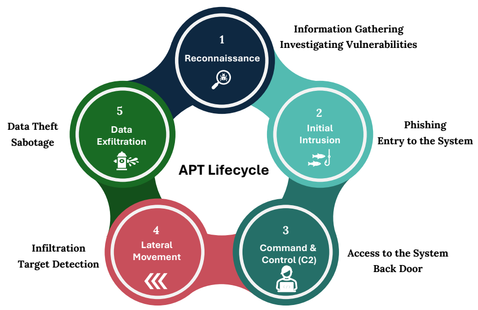

Figure 2: The five-stage Lifecycle of an APT

3.1.3 Tactics, Techniques, and Procedures (TTPs)

TTPs are shown in Figure 3 and are the framework used to classify the behavior of an APT attack. It can be defined as follows [27]:

- **Tactics**: Used to define the goal of the attack, such as gaining access to a system.
- **Techniques**: Refers to the technique used to achieve this goal. An example would be DLL injection.
- **Procedures**: Refers to the way the attack is implemented, such as sending a special email.

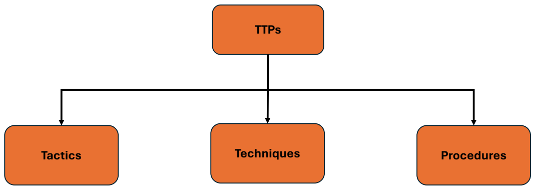

Figure 3: The Hierarchical Structure of TTPs

3.1.4 APT vs. Traditional Attacks

APTs and traditional attacks differ from each other in many ways, such as target, tactics, duration, and these differences are shown in Table 2. Traditional attacks aim to cause general damage and aim for quick gain, while APTs are long-term and professional attacks (usually state-sponsored) [26]. Traditional attacks identify weak systems by simultaneously attacking many targets, while APTs are more target-oriented and the attack process is carried out in a long and sneaky way [28].

Table 2: Comparative Characteristics: Traditional Attacks vs. Advanced Persistent Threats (APT)

Attribute	**Traditional Attack**	**Advanced Persistent Threat (APT)**
Target	Broad or Random	Highly Specific
Duration	Short-lived	Prolonged (Months or Years)
Entry Vector	Known Exploits	Custom Zero-Days, Spear Phishing
Goal	Financial Gain, Disruption	Espionage, Strategic Access
Tools Used	Commodity Malware	Tailored, Multi-Stage Toolkits

3.2 6G Wireless Networks

3.2.1 Architectural Foundations of 6G

6G is the new generation of wireless communication paradigm that emerges with the integration of advanced physical technologies and software-defined network solutions [29]. In order to provide uninterrupted communication in the 6G architecture, it is a heterogeneous structure (Ultra-Dense Heterogeneous Networks) that combines three basic layers: terrestrial, aerial, and satellite [30]. In order to reach a data rate of more than 1 Tbps, technologies such as Terahertz (THz) communication and Visible Light Communication (VLC) are used [31]. In addition, power-sensitive technologies such as Reconfigurable Smart Surfaces (RIS) and Software-Defined Metasurfaces (SDM) are used to reduce latency [32].

The 6G networks with heterogeneous architecture shown in Figure 4 try to reduce computational loads with edge devices and AI-based systems [33]. While 5G architectures use a centralized system, in the 6G architecture, thanks to decentralization, network slices can be optimized autonomously via AI-based engines. However, despite these advantages, the heterogeneous and decentralized architecture offers a large attack surface and is exposed to cyberattack threats [33].

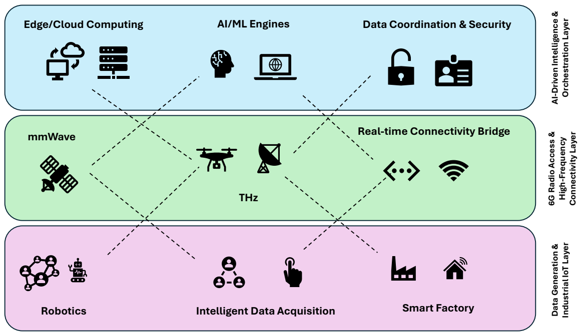

Figure 4: 6G Architectural Pillars and Deployment Layers

3.2.2 Key Features of 6G

Table 3 shows a feature comparison of 5G and 6G, and as can be seen, 6G (key features) is superior in every aspect. 6G is expected to work in harmony with real-time holography and trusted autonomous systems once it is available for daily use [34]. 6G relies on AI-powered protocols and advanced infrastructure capabilities to meet these demands [35].

Table 3: Comparison of Key Features Between 5G and 6G Wireless Networks

Feature	**5G**	**6G**
Latency	Approximately 1 millisecond	Less than or equal to 0.1 milliseconds
Data Rate	Up to 20 Gbps	At least 1 Tbps
Frequency Range	Sub-6 GHz and millimeter wave	Sub-Terahertz (Sub-THz) and Visible Light Communication (VLC)
Architecture	Centralized network control	Distributed and AI-powered network architecture
Security	Add-on security features	Built-in, intent-aware security mechanisms

3.2.3 Vulnerabilities and Threat Landscape in 6G

Despite the high speed and wide infrastructure opportunities they offer, 6G networks also carry risks such as misconfiguration and hostile exploitation due to AI-based control logic and network software such as SDN/NFV [36]. If vertical slicing and segmentation operations in networks do not work correctly, they become vulnerable to lateral attacks (sourced by APT's etc.) [37].

Possible potential attacks that may occur in 6G are shown in Figure 5. Attack types can range from physical layer compression to manipulation. Another potential danger is that the AI mechanisms responsible for 6G orchestration are vulnerable to attack and data leakage in cases where RIS and THz communication channels are not properly set [38]

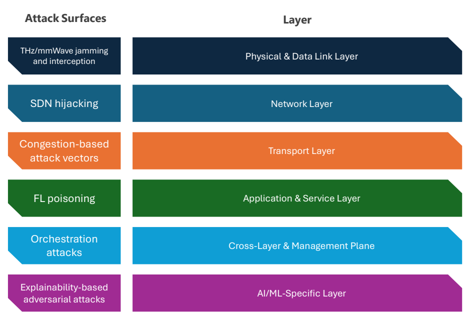

Figure 5: Illustration of Potential Attack Surfaces Across the Hierarchical 6G Network

3.2.4 6G-Specific Challenges for APT Detection

APTs are expected to threaten the rapid lateral movements that 6G will bring to our lives [39]. In addition, behavioral detection becomes more complex for traditional signature-based intrusion detection systems (IDS) due to architectural features such as encrypted layers and dynamic topologies [40].

Another challenge for 6G networks is the scarcity of APT datasets, which makes it difficult to train security models [41]. Another challenge is the fragmented nature of source logs, as this limits the correlation between layers that can be used in APT detection [42].

3.2.5 Research Trends Integrating 6G and AI for Security

Literature studies investigate the use of FL at edges to detect attacks while also concerning privacy [43]. In addition, XAI methods for decision-making mechanisms are another frequently investigated method [44]. Beyond these, mapping TTPs and analyzing logs with LLM's based systems are promising [45]. However, since edge devices are resource-constrained and storage-limited devices, these limitations should be taken into consideration when deploying LLM's at edges, and strategies such as model distillation should be applied [46].

3.3 LLM's for APT Detection in 6G Networks

3.3.1 Overview of LLM Architectures and Security-Oriented Specializations

Developed on Transformer architecture, LLM's have made a great breakthrough in the field of AI, and these models provide representation learning by making sense of the context [47]. In other words, these LLM's indicate the ability to understand the meanings of words in context beyond their dictionary meanings. Thanks to this ability, they achieve great success in natural language understanding (NLU) and generation (NLG) tasks [48].

This technological development (LLMs) has begun to be used in many areas, especially in cybersecurity, and the evolution of LLMs in cybersecurity use is shown in Figure 6. These areas of use can be examined under three main headings [49, 50]:

- **General-Purpose**: LLM models that can be trained with large text collections and used for various purposes.
- **Domain-Specific**: LLM models trained (fine-tuned) using purpose-oriented cybersecurity data.
- **Emerging Techniques**: These are methods that make LLM's models lightweight and specific to their intended use.

General-purpose models such as BERT (2018), GPT-2 (2019) have shown success in text classification and question-answer tasks, and with the customization of these models, domain-specific models such as SecBERT (2020), CyBERT (2021) have been developed and started to be used in special tasks such as malware detection (such as APT). And studies on emerging techniques continue to increase the performance of these models.

3.3.2 Applications in Cybersecurity and APT Detection

LLMs have been used in cybersecurity for multi-APT detection and response, returning based on attack type [51]. LLM's features, such as contextual reasoning and linguistic understanding, make it particularly suitable for APTs with multi-stage attacks [52]. Figure 7 shows a tree structure for LLM application areas in APT detection. As can be seen from the figure, LLMs are versatile in the cybersecurity context [53, 54, 55]:

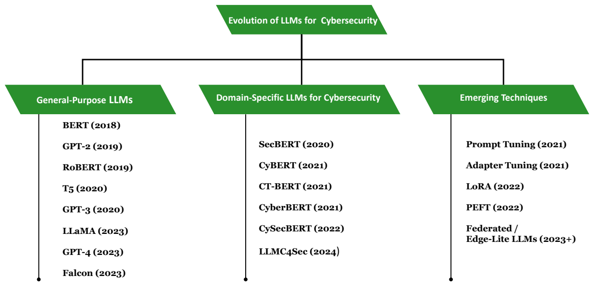

Figure 6: Evolution of Large Language Models (LLMs) for Cybersecurity

- **Threat Intelligence**: LLM's models can extract TTPs using open-source data such as threat reports.
- **APT Behavior Modeling**: Logs and lineage data can be used to semantically interpret multi-stage APTs.
- **Anomaly Detection**: LLMs context-aware feature can detect anomalous behavior (network and system logs).
- **Alert Triage and Incident Response**: Natural language summarization translates alerts into insights to extract meaningful information (helpful for analysts).
- **TTP Alignment**: LLMs fine-tuning can map hostile behaviors for low-fire systems to MITRE ATT&CK stages.

In dynamic networks (especially 6G networks), the GPT family provides situational awareness in complex architectures where traditional detection methods can struggle.

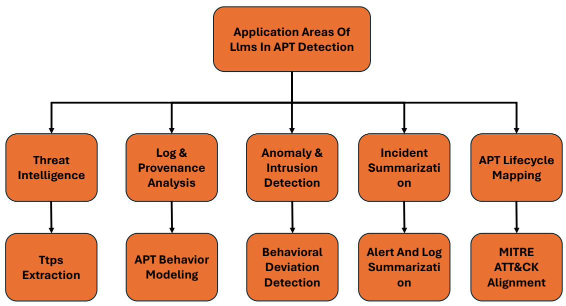

Figure 7: Application Areas of LLMs in APT Detection

3.3.3 LLM Integration Challenges in 6G Edge Environments

Edge devices positioned close to the data source have advantages such as low latency and low bandwidth usage, but also disadvantages such as heterogeneous structure and limited processing power [56]. For this reason, problems arise due to these limitations when LLMs are deployed on 6G-based edge devices. Figure 8 summarizes these limitations and possible solutions [1, 2]:

- **Resource Constraints**: Since LLMs require high processing power and storage, their implementation on resource-constrained 6G edge devices (RAM & Compute Energy Efficiency) is one of the major problems that can be encountered.

- **Latency Constraints**: Edge devices, which are expected to offer a low latency advantage due to being positioned close to the data source, may lose this advantage due to the high computational time of LLMs.
- **Privacy & Compliance**: Sensitive data such as biometrics must take into account some privacy concerns when processed on edge devices [56, 57].

The methods to solve these challenges can be summarized as follows [58, 59]:

- **Compression**: LLMs can be downgraded to lower versions to reduce memory and processing load.
- **Knowledge Distillation**: Information obtained from large models can be transferred to smaller models to minimize performance loss.
- **Federated/Split Inference**: Both privacy and efficiency can be increased by distributing the components of the LLM model to different edge nodes and processing them.

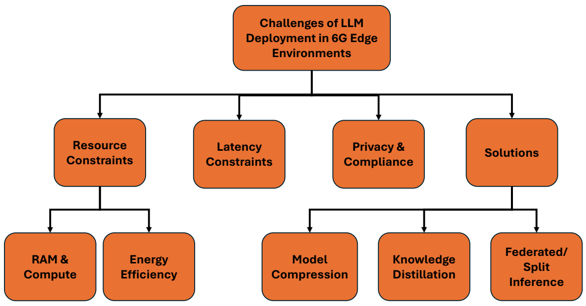

Figure 8: Challenges of LLM Deployment in 6G Edge Environments

3.3.4 APT Detection-Specific Benefits of LLMs in 6G Context

LLM models have great potential in APT threat reduction studies in 6G networks, which are expected to be used in the near future. This potential stems from the success of LLM models in establishing semantic correlations between data types and their ability to analyze the attack lifecycle as a whole [60]. The contributions of LLM models for APT detection in 6G networks can be generalized as follows [61, 62]:

- **Cross-Layer Fusion**: It can be used in the detection of multi-vector attacks by combining log records from the control and user planes and the cloud layers.
- **Lifecycle Prediction**: LLM models can use past attack data to predict the next step in the APT kill-chain.
- **Semantic Generalization**: LLM models can capture attacks in a contextual manner in encrypted and hidden attack situations where traditional systems are inadequate.

Figure 9 shows the contributions of LLM models for layers. As can be seen from the figure, LLM models can perform multi-layered threat modeling by performing detection not only at the packet level but also at various levels of the network.

In addition, LLM models not only interpret the behavior of APT attacks but also provide insight into the attack lifecycle phases and response mechanisms. Figure 10 shows the tasks that LLMs undertake in the APT kill chain model.

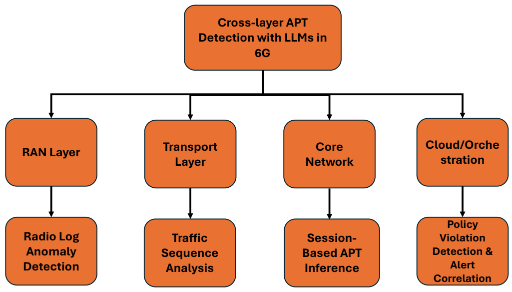

Figure 9: Cross-layer APT Detection with LLMs in 6G

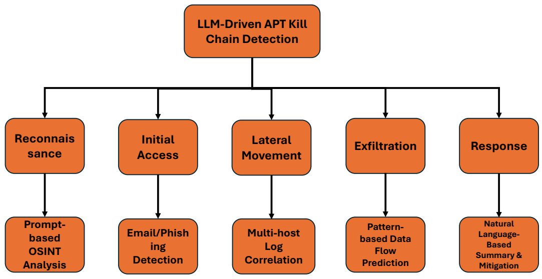

Figure 10: LLM-Driven APT Kill Chain Detection

3.3.5 Proposed Taxonomy: LLM-Driven APT Detection in 6G

This paper aims to classify LLM-based APT detection approaches in 6G networks and provide a comprehensive taxonomy. Figure 11 shows this taxonomy and can be summarized in five dimensions:

- **Input Modalities**: LLM models can be fed from various data sources such as logs and PCAP.
- **Detection Granularity**: LLMs can perform APT detection at different levels, such as single-packet analysis and session-based modeling.
- **LLM Techniques**: LLM models can be trained in various ways (prompt tuning, etc.) according to different scenarios.
- **Deployment Models**: LLMs can be deployed on different platform environments, such as cloud computing and edge computing.
- **Threat Lifecycle Phase**: LLM models can provide analysis and interventions at various stages of the APT kill chain.

4 REVIEW METHODOLOGY

This section presents the methodologies employed in the systematic review of LLM-driven APT detection approaches within 6G wireless networks and subsequently outlines the formulated research questions.

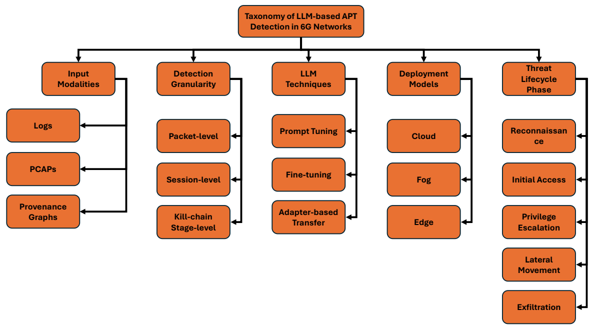

Figure 11: LLM-Based APT Detection Taxonomy in 6G Networks

4.1 Papers Collection

Since LLM-focused APT detection approaches in 6G wireless networks are a very current topic, we targeted the years 2018-2025 (current) to collect the relevant current literature studies. The following keywords were used to identify the studies related to the research topic:

1. `[(LLM)` — (LargeLanguageModel)] & [(APT) —— (AdvancedPersistentThreat)]—
2. `[(6G)` — (WirelessNetworks)] & [(LLM) —— (APTDetection)] & [(Edge) —— (Cross-LayerSecurity)]—
3. `[(CyberThreatIntelligence)` — (ProvenanceLogs)] & [(LLM) —— (APT)] & [(6G)]—
4. `[(LLM)]` — [(APT)] —— [(6G)]—

Figure 12 shows the collection and filtering process of the articles examined in this study. The steps carried out for this process are aimed at providing a comprehensive and structured analysis by following Kitchenham's Systematic Literature Review (SLR) and Petersen's Systematic Mapping Study (SMS) approaches [7, 8]. The steps summarizing this process are explained below:

- **Identification**: Known major academic literature sources (IEEE, ACM, Elsevier, Springer), technical reports, book chapters, and reference lists were scanned.
- **Screening**: Duplicate documents were removed, and the number of papers decreased to 126.
- **Eligibility**: Papers collected by our expert authors were analyzed, and only quality and scope-compliant papers were selected (the number of papers decreased to 120).
- **Included**: Additional relevant studies were added using the backward and forward snowball method, and the paper set was determined as 142 [63].

4.2 Research Questions

The research questions used in the systematic review and the section in which they are examined are shown in Table 4. Using these research questions, the current literature is examined and analyzed.

5 ANALYSIS

In this section, we discuss how we addressed the research questions in this study.

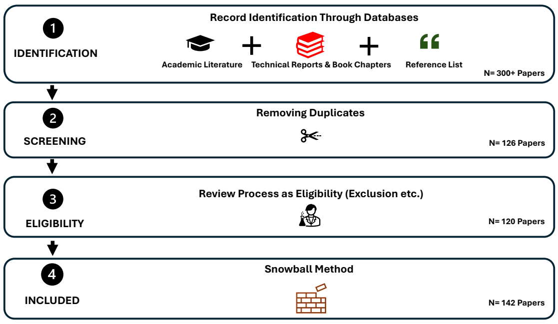

Figure 12: The Paper Collection Process

Table 4: Summary of Research Questions (RQs), Motivations, and Corresponding Sections

NO	RQ	Motivation	Section
1	How can LLMs semantically use fragmented provenance data from 6G sources in APT detection?	The aim of this RQ is to investigate how fragmented resource records can be attributed by LLMs in 6G.	5.1
2	What are the limitations of encrypted 6G channels in APT detection and what features of LLM can solve this problem?	The aim of this RQ is to examine the potential challenges posed by 6G networks in APT detection and how LLMs can solve them.	5.2
3	What are the resource constraints in LLMs deployment to 6G edge devices and what techniques can be used to mitigate these limitations?	The aim of this RQ is to explore which compression strategies can be applied when deploying LLMs models to 6G edge devices.	5.3
4	What are the datasets and modeling methods available for LLM-driven APT detection approaches within 6G wireless networks?	The aim of this RQ is to investigate suitable datasets and dataset generation methods for LLM-focused APT detection approaches in 6G wireless networks.	5.4
5	Where are existing LLM-based APT studies published and do they support reproducibility?	The aim of this RQ is to evaluate the reproducibility of the dataset and model usability of the reviewed studies.	5.5

5.1 Semantic Correlation of Fragmented Provenance Logs in 6G (RQ1)

6G networks contain a lot of fragmented lineage data due to their heterogeneous structure (edge, cloud, etc.), and this data is distributed and inconsistent, which causes difficulties for security analysts and attack systems in APT detection [64]. One example of these difficulties is that rule-based and statistical detection methods fail to capture the nuanced context required for attack detection [65, 66]. Recent studies have focused on LLM-based methods to semantically combine fragmented lineage data and provide context-aware correlation. Figure 13 shows how fragmented lineage data can be semantically associated with LLM-enabled systems in a multi-layered manner. LLMs offer a promising solution by generating consistent security narratives by syntactically and temporally handling various records (such as security logs) [67, 68, 69, 70, 71].

Recent findings have shown that LLM's models can effectively utilize many different sources, such as audit logs [67], IDS alerts [4], CTI reports [72], and even static code artifacts [73]. Models that transform low-level source sequences into textual formats, such as APT-LLM [67], GENTTP [72], and LLMeLog [69], have been developed to reflect the system behavior semantics of models such as BERT or RoBERTa. Frameworks based on multitasking instructions and thought chains, such as SEVENLLM [74] and AnomalyGen [75], have been proposed for reasoning in data-scarce environments. For enrichment techniques, the literature includes studies such as retrieval-augmented generation [68], clustering embedding [76, 69], and ATT&CK alignment via request templates [77].

Many frameworks have been proposed that support fragmented logs with deep reasoning by capturing temporal, causal, and entity-level relationships and that resort to graph-based modeling. SHIELD [68], MultiKG [78], and MAD-LLM [79] frameworks use source graphs that encode dependencies of edges and represent system events at nodes. Other works such as AURORA [80] and DroidTTP [71] reconstruct attack sequences by applying classical planning and LLM. Works such as LocalIntel [81] and MCM-LLAMA [82] prefer dynamic association of SOC information and external alerts, while works such as LUNAR [76] and AnomalyGen [75] prefer association with CTI corpora. For high-level reasoning and explanation generation, these works resort to semantically annotated graph-based modeling.

Despite all these developments, there are still limitations that remain to be addressed. SEV-

ENLLM [74] and SHIELD [68] frameworks use organized and synthetic logs, but this does not fully reflect the dynamic, heterogeneous nature of 6G. Another point to note is that mitigation strategies such as hybrid verification [67] and instruction fine-tuning [74] are rarely applicable to edge contexts. In addition, LLMs' high processing power and storage requirements make their application in 6G edge nodes a serious challenge, and therefore, the need for lightweight alternative methods such as TinyLM agents [83] or MoE-based distributed inference [79] is increasing. For future research, areas such as cross-layer lineage fusion, real-time semantic timeline reconstruction, and hallucination-aware causal modeling [73] stand out as a research gap.

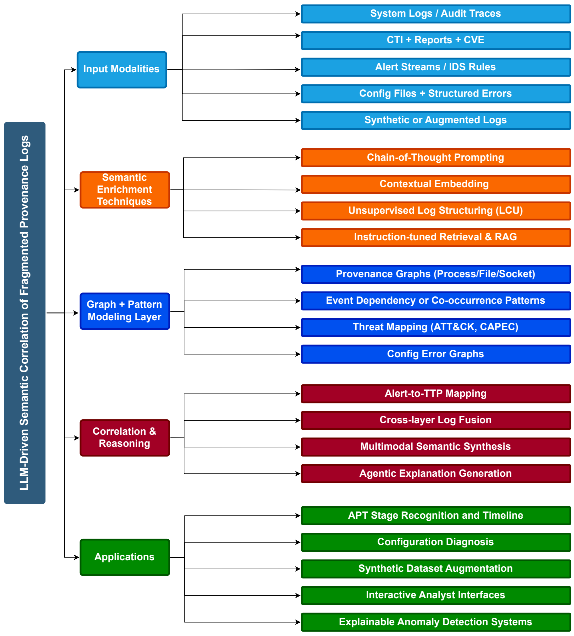

Figure 13: RQ1 Taxonomy: LLM-based Semantic Correlation of Fragmented Provenance Data Across Heterogeneous 6G Sources

5.2 Limitations of Encrypted 6G Channels and LLM-Driven Solutions (RQ2)

The widespread use of some communication protocols, such as DNS-over-HTTPS (DoH) and end-to-end encrypted tunnels, in the transition to 6G wireless networks has made great contributions to security and user privacy. In addition to these contributions, it also brings disadvantages like blind spots, such as traffic semantics obscurity for AI-supported detection systems. Figure 14 shows a taxonomy of LLM-focused solutions offered to address the challenges, limitations, and risks of 6G networks due to encrypted channels.

5.2.1 Technical Limitations Imposed by Encryption

Encrypted 6G traffic channels limit the visibility of attack surfaces due to the techniques used (such as DoH). Recent studies have shown that advanced DL models fail to detect malicious traffic because semantic payloads become ambiguous while being encrypted [84]. In addition, advanced attack methods such as APT try to avoid detection by using encrypted channels such as DoH and embedding the C2 infrastructure in HTTPS payloads [85]. Edge-based data isolation, whose main purpose is privacy, prevents correlation (temporal and spatial) between devices. For example, since fragmented traffic logs are produced in UAV-based 6G networks, anomaly monitoring becomes very difficult [86]

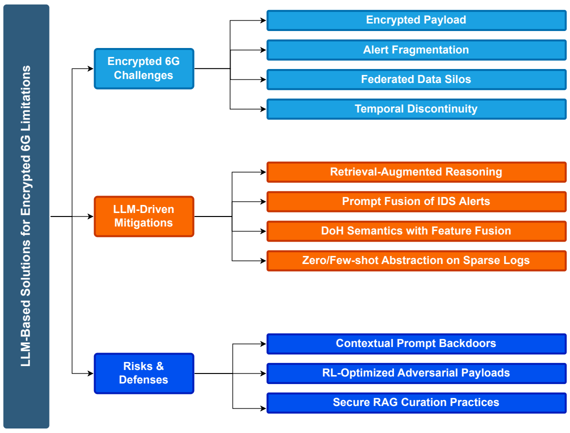

Figure 14: RQ2 Taxonomy: LLM-Based Solutions for Encrypted 6G Limitations

5.2.2 LLM-Driven Mechanisms to Address These Gaps

To overcome all these limitations, LLMs are promising by making meaningful inferences with their capabilities in semantic reasoning and contextual abstraction.

A recent study, APTSniffer, is a framework that detects APTs in encrypted channels by converting flow features into textual prompts [87]. The results confirm that the framework is successful with a 97% F1 score. Another study, MAD-LLM, is a framework that reconstructs APT chains by semantically collecting them through LLMs despite fragmented IDS alerts and encryption at the network layer [79]

Some malware (such as DoHunter, Godlua) are difficult to detect by detection systems because they use encrypted channels, so researchers track some technical features of the traffic, such as timing, length, and target domain structure, in addition to raw data with LLM models [85].

5.2.3 Emerging Challenges and Threats

Although using LLM models in encrypted channels is a promising solution, it is important to consider LLM-based vulnerabilities. One of these vulnerabilities is that LLM behavior can be manipulated by hostile requests and poisoned rollbacks. Studies have confirmed that fine-tuned LLM models based on RL can generate malicious traffic [84]. Furthermore, literature confirms that LLM models inject hidden logic into LLM models that are activated by benign triggers in encrypted channels [88]

In conclusion, while LLM models provide an advantage, such as semantic visibility for attack detection in encrypted channels, they also inherently introduce attack surfaces.

Table 5 summarizes the main limitations of encrypted 6G environments in light of the current literature reviewed.

Table 5: Mapping Encrypted 6G Challenges to LLM-Driven Solutions

Work	**Limitation**	**LLM-Based Technique**
Xu et al. [87]	Payload Obfuscation	Retrieval-Augmented Inference
Du et al. [79]	Alert Fragmentation	Multi-stage Reasoning via Prompt Engineering
Diao et al. [85]	Covert DoH C2 Channels	LLM + Expert Features for Tunnel Detection
Cheng et al. [89]	Contextual Reasoning in Sparse Logs	Log Fusion and Interpretation via Few-shot Learning
Sun et al. [84]	LLM Model Poisoning via Traffic	Adversarial Sample Generation with Reinforcement Learning (RL)
Liu et al. [88]	Contextual Logic Corruption	In-context Backdoor Prompt Manipulation

5.3 Deploying LLMs at the Edge: Constraints and Optimization Techniques (RQ3)

Despite the advantages of high speed and low latency offered by 6G networks, they consist of many different distributed nodes and heterogeneous structures, such as edge devices. Therefore, LLM models to be used for security, privacy, and context-adaptive smart applications should also take into account the major computational, architectural, and security-related challenges when deployed in 6G networks. This research question (RQ3) examines optimization techniques for edge scenarios by categorizing these constraints. Figure 15 shows edge-oriented LLM optimization strategies.

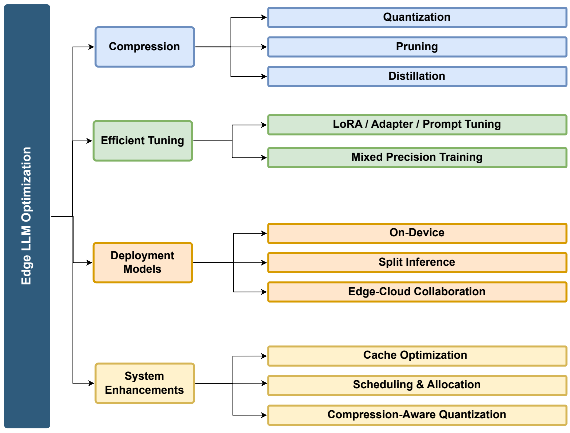

Figure 15: RQ3 Taxonomy: Edge-Oriented LLM Optimization Strategies

Resource Constraints in Edge Environments: Edge devices (IoT, smartphones, etc.) consist of devices with limited processing power and storage capabilities, and even very small LLM models require more than 7GB RAM, which is usually beyond the capacity of edge computing devices [90]. Furthermore, due to the nature of LLMs (autoregressive), sequential token generation may cause latency bottlenecks [91].

Security and Fairness Considerations: Since LLM models deployed on edge nodes typically handle user data, privacy concerns may arise if this data is compromised [3]. Additionally, recent studies have reported that compression techniques used to make LLM models lighter for edge nodes may increase bias against underrepresented groups [92]. Therefore, the issue of fairness and reliability in compression cases is an open research gap.

Model Compression Techniques: These are techniques used to reduce the memory and processing load of LLM models, and one of the most popular methods is quantization. In this method, the weights (such as FP32) are converted to lower bit representations (INT8 or FP4) to reduce the model size and optimize the hardware speed [93, 94]. Another method where the distribution is optimized is pruning, and in this method, the weights are rescaled before quantization [90]. Distillation and low-rank approximation methods aim to provide additional performance gains on the inference quality [92, 93].

Parameter-Efficient Fine-Tuning (PEFT): Fine-tuning of LLM models cannot be done at edge nodes, and therefore PEFT methods (such as LoRA, Adapters, and Prompt-Tuning) apply them to local tasks by updating a subset of the models' parameters [95]. Another recent research introduces a new collaborative training model that optimizes fine-tuning of early layers at the edge device (mobile) and deep layers at the edge server [96]. In this model, the aim is to reduce communication and energy costs while keeping the personalized performance constant.

Collaborative and Split Inference: Another rational approach is to distribute the LLM model across the device-edge-cloud heterogeneity to achieve the balance of performance and resource utilization. Yang et al. [97] propose a structure in which the cloud and edge are jointly used to offload LLM inference with a UCB-based scheduler. The results show that the energy usage is halved and the efficiency is doubled. Another approach is to share the converter layers among the heterogeneous edge devices using the matching theory [98]

System-Level Runtime Optimizations: Another method of increasing efficiency is runtime and architecture-based approaches:

- **KV-cache compression**: It is the process of reorganizing memory to save RAM on edge devices [94]
- **Contextual sparsity and batch-aware scheduling**: In contextual sparsity, the process of reducing the processing load by looking only at the important tokens of the model, while in batch-aware scheduling, the process of running multiple tasks in the most efficient order without blocking each other [96, 94].
- **Speculative decoding**: It is the process of predicting multiple tokens at the same time, and the aim is to reduce the autoregressive latency bottleneck [91].

Table 6 summarizes the main optimization techniques developed against resource constraints on edge devices, their usage scenarios, and the cost/benefit balances they bring.

Table 6: Optimization Techniques for Deploying LLMs under Edge Constraints

Reference	Technique	Constraint Addressed	Use Case	Trade-off
Wang et al. [93]	INT4/INT8 Quantization	Memory, Compute	Mobile Edge Inference	↓Accuracy, ↑Speed
Wang et al. [90]	Compression-aware Quantization	Memory, Latency	Smartphone + AI Assistant	Low overhead
Qin et al. [95]	LoRA / Adapters	Fine-Tuning cost, Storage	Personalized edge assistants	↓Flexibility, ↑Privacy
Yang et al. [97]	PerLLM Scheduler	QoS-aware Scheduling	Edge-Cloud Mixed Load	↑Efficiency, ↑Throughput
Zhao et al. [91]	Token Parallel Decoding	Latency (token gen)	Edge-terminal co-inference	Complex sync
Picano et al. [98]	Matching-based Layer Placement	Device Heterogeneity	Heterogeneous Edge Inference	↑Accuracy

5.3.1 Federated or Distributed Training of LLMs at the Edge

With the proliferation of user-centric applications, it is expected that deployment strategies of LLM models will be developed on edge devices as well [88]. Cloud-based systems cause additional latency and bandwidth loads in 6G environments compared to edge computing [1]. Therefore, Federated Learning (FL) and distributed tuning paradigms can be used to reduce these loads by processing data on edge devices [1].

Motivation for Federated Edge Training:

Compared to traditional cloud-based systems, training on edge devices provides improvements in the following limitations [3, 99, 91]:

- **Privacy**: Since sensitive data, such as biometric data, is processed on edge devices, there are fewer privacy concerns than cloud systems that use central servers.
- **Latency**: Since data is processed close to the data source, there is less latency than cloud-based systems.
- **Bandwidth**: Since cloud-based systems are used only for operations that require large processing power, unnecessary communication bandwidth is not used.

Federated Fine-Tuning Techniques for LLMs:

In constrained environments (such as communication and computation), LLM model personalization schemes can be summarized as follows:

- **Parameter efficient**: In LoRA-based work, low-rank matrices are fine-tuned among clients to reduce transmission volume [96].

- **Split Federation Learning**: Qu et al. [99] propose to train the first layers of the model on the device and optimize the deep layers on edge nodes in their proposed framework called Mobile Edge Intelligence (MEI).

- **Inter-device gradient fusion**: In order to dynamically balance the update frequency and energy budgets, distributed scheduling algorithms are proposed in [91].

Challenges in Federated LLM Training:

Despite the advantages of low latency and low bandwidth overhead, federated training at the edge also suffers from statistical heterogeneity [93], system heterogeneity [100], and security risks [3]. To overcome these challenges, recent research focuses on model-system co-design based techniques. These techniques include adaptive aggregation [97] where clients are weighted according to their trust scores, compression-aware updates [90] where updates are sparse before transmission, and energy-aware scheduling [94] where the training frequency is dynamically adjusted to preserve battery and network life. Table 7 provides a comparative overview of federated and distributed education strategies.

Table 7: Federated and Distributed LLM Training: Constraints and LLM-Based Trade-offs

Work	**Constraint Addressed**	**LLM-Based Technique / Trade-off**
Liu et al. [96]	Bandwidth, Memory	LoRA-Based FL for Lightweight Personalization (Lower Global Accuracy)
Qu et al. [99]	Compute Offloading	Split Learning (MEI4LLM) with Multi-layer Collaboration (Sync Overhead)
Zhao et al. [91]	Latency, Energy	Parallel Token Learning at Edge-Terminals (Complex Scheduling)
Qin et al. [95]	Fine-tuning Privacy	Federated Prompting + PEFT (Bias Sensitivity in User-Tuning)
Yang et al. [97]	Device Reliability	Trust-Aware Aggregation in FL (Risk of Model Divergence)

5.4 Datasets and Modeling Techniques for LLM-Driven APT Detection (RQ4)

The quality of the datasets to be used to train models in LLM-based APT detection in 6G networks directly affects the success rate. The datasets created as a result of examining 32 different studies and the results of the systematic and taxonomy study on modeling techniques are examined in this subsection. Figure 16 shows this taxonomy and its subsections.

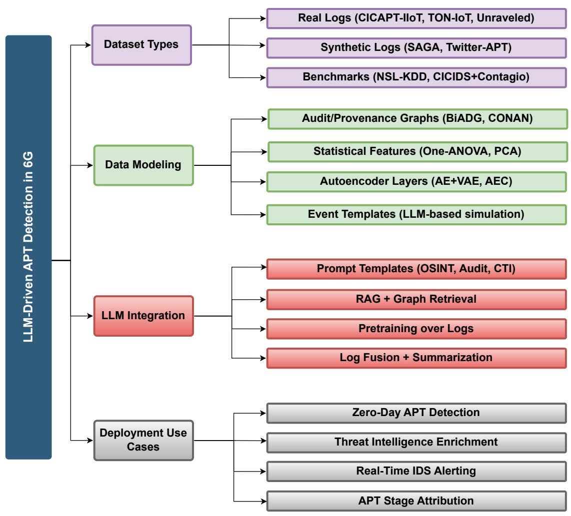

Figure 16: Taxonomy of Dataset–Model–LLM Alignment in 6G APT Detection Pipelines

Dataset Types for LLM-Based APT Detection: When the literature is examined, it is seen that APT datasets can be examined under three main headings:

- **Semi-Synthetic Datasets**: Semi-synthetic datasets that model APT attacks are as follows: (i) Unraveled dataset [101], which combines real cloud infrastructure logs and simulated APT stages, and (ii) edge-based CICAPT-IIoT dataset [102], which includes UAV and ICS smart environment logs.

- **Synthetic and Augmented Logs**: Synthetic datasets that model APT attacks are as follows: (i) SAGA [103], which consists of audit logs compatible with the ATT&CK matrix, (ii) Twitter-APT [104], which is created by applying LLM's to OSINT-based threats, (iii) a dataset where labeled attacks are created using pcap filters, log records and IDS simulation [105].

- **Merged Benchmark Corpora**: For APT detection, datasets that are based on a real organization's network traffic and model attacks such as trojans and spyware can be used [106].

NSL-KDD or CICIDS data outside these categories are now outdated and fail to model real APTs (stealth lateral movement or long-term dormancy strategies) [107, 108]

Data Modeling Techniques and Representations: Data modeling strategies that can be used to combine 6G network data with LLM models can be summarized as follows:

- **Behavioral Graph Profiling**: In this modeling method, BiADG and MIG models are obtained by applying Graph Convolutional Network (GCN) on IP flow graphs and behavior patterns [109, 110]. In addition, there is the CONAN model that provides low-latency matching for APT stages using a Finite State Machine (FSM) [111].

- **Statistical + Feature Engineering Pipelines**: As an example of these strategies, two separate studies that apply preprocessing such as One-ANOVA based cleaning, decomposition, and boosting by synthetic generation [106, 105] can be given as examples.

- **Multi-Stage Autoencoders**: In APTSID [106], where this strategy is applied, standard and variational autoencoders are combined with statistical feature selection to achieve high accuracy anomaly detection.

- **ML + Expert System Hybrids**: In the CDT system [105], where this technique is used, an attack detection prediction is taken with an ML model and transmitted to the rules used by systems such as SMORT.

Table 8 shows the comparison of datasets and modeling techniques in LLM-Driven APT detection studies

LLM Integration Strategies: These are the methods used when integrating LLM models into various systems, and the main purpose is to enable LLM models to be used with various data.

- **Prompt Templates + Simulation**: LLM prompts are the methods used to generate attack data and multiply training data, and SAGA and CyExec are two examples of this in academia [103, 112].

- **OSINT + NER Pipelines**: Although LLM models are successful in detecting threats in open source articles, fine-tuning is required for small details. Shafee et al.[104] tries to find threats from open source information with LLM.

- **Fusion Architectures**: In these methods, after the data is processed with other models and made meaningful, it is given to the LLM model to process, and thus it is expected that LLM will perform a more successful analysis. Models such as AE+VAE and AE-CNN use this method [106]

Table 8: Comparison of Datasets and Modeling Techniques in LLM-Driven APT Detection Studies

Reference	Dataset	Modeling Technique	LLM Use	Key Insight
Huang et al. [103]	SAGA (Synthetic)	Prompt-Based Log Generation	Training Input	ATT&CK-aligned, synthetic audit logs for APT stages
Neuschmied et al. [106]	CICIDS + Contagio	AE + VAE Stack	Feature Compression	Multi-stage anomaly detection with zero-day support
Al-Aamri et al. [105]	Custom Logs	CDT + SNORT Rule Feed	Manual LLM Friendly	Time-series + journaling logs with rule generation
Ghiasvand et al. [102]	CICAPT-IIoT	Provenance + Network Flow Fusion	LLM-Graph Possible	Audit trails + flow logs for IIoT APT detection
Shafee et al. [104]	Twitter Corpus	OSINT Classification + NER	NER and Prompt Evaluation	LLMs need domain adaptation for threat-level NER
Xuan et al. [109]	BiADG	Behavioral GCN + LSTM	Graph-to-LLM Potential	IP-node behavior modeling over graph structures
Oleiwi et al. [107]	NSL / CICIDS / UNSW	Meta-Model Voting Ensemble	Pre-LLM Classifier Layer	Traditional ML stack for high-precision filtering

5.5 Reproducibility and Publication Trends in LLM-Based APT Studies (RQ5)

This research question questions the reproducibility and other statistical information of LLM-based APT detection studies. In order to provide a comprehensive assessment of the 142 recent studies utilized throughout the paper, we have classified all papers in **Appendix A** according to *Code Availability*, *Dataset Evaluation*, *Protocol Venue/Platform*, and *Year*. The description of these features and the resulting statistical information are as follows:

Code Availability: This feature was used to classify studies according to their reproducibility. Figure 17 shows how many percentage of the studies shared their source code (YES/NO), and on which platform (Github, etc.) they were published. As can be seen from the figure, only a very small portion of the examined studies shared their source code, while most of their code was published on the GitHub platform.

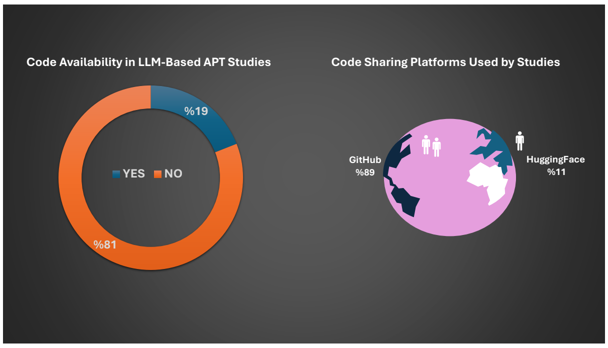

Figure 17: The Percentage of LLM-based APT Studies that Shared Source Code (YES/NO) and the Platforms where the Code was Hosted

Dataset: This column was used to measure the diversity of datasets used in the studies and to determine how many of them used real-world data. Figure 18 shows the percentage of datasets shared by year and the percentage of articles using synthetic-public datasets. The results confirm that datasets used in APT detection studies tend to be shared and that the most used dataset is synthetic dataset.

Evaluation Protocol: This column is to evaluate the level of empirical validity of the reviewed articles based on whether they use robust protocols such as cross-validation. Figure 19 shows the frequency of the protocols used.

Venue / Platform: This column examines the publication quality and field spread by examining the venue/platforms and types (conference/journal) where the reviewed studies were published. Figure 20 shows a summary of this statistic.

Year: The last column shares the publication dates of the reviewed studies and evaluates the increase in LLM-focused APT papers as we move towards the 6G wireless networks era.

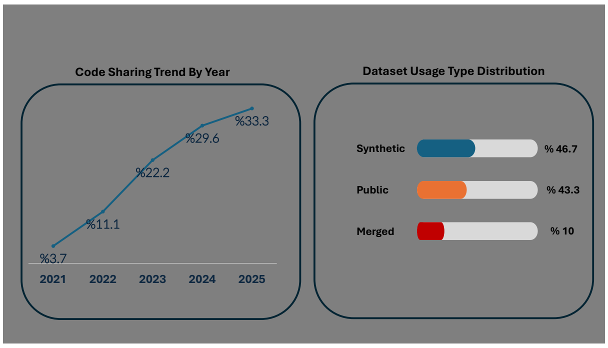

Figure 18: The Trend of Dataset Usage over the Years and Most Commonly used Datasets in LLM-based APT Studies

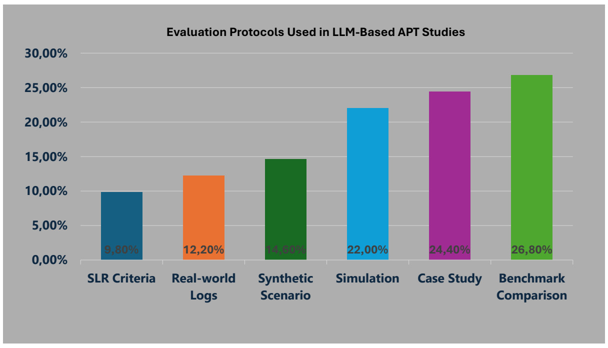

Figure 19: Evaluation Protocols Used in LLM-Based APT Studies

Figure 21 shows the change in LLM-focused APT papers by year.

6 OPEN CHALLENGES AND FUTURE DIRECTIONS

As the use of 6G networks and LLM deployments in 6G becomes widespread, many research gaps and new open challenges to be solved will emerge for researchers. These challenges include architecture and security issues, and we discuss the open challenges, a taxonomy of which is given in Figure 22, in this section.

Semantic-Aware Reasoning and Limited Contextual Memory: LLMs are promising for APT detection in 6G networks with their high performance in understanding causal relationships and threat contexts using data such as system logs and audit trails [107]. However, LLM models have limited performance in long-term and fragmented event sequences because their architectures offer limited window sizes and context management. Therefore, it makes detection difficult in multi-stage APTs with long processes such as infiltration and reconnaissance.

Future Directions: Future researchers can overcome these limitations by focusing on the integration of memory modules and hierarchical memory structures. These structures make it easier for the model to learn long-term correlations between events. An example of this is the modeling of the relationship between system input and data exfiltration behavior to understand the holistic behavior of an attack. In addition, models that can provide transformer-graph synergy, such as GNNs, can effectively establish topological or temporal relationships between events [18]. Thus, event traces can be modeled over a graph structure, enabling LLM models to learn event dependencies in a scalable manner.

Real-Time Processing Under Edge Constraints: 6G lines provide great advantages for time-constrained scenarios such as autonomous vehicles by offering high speed and low latency [64]. In particular, since edge devices in their heterogeneous structure bring the processing

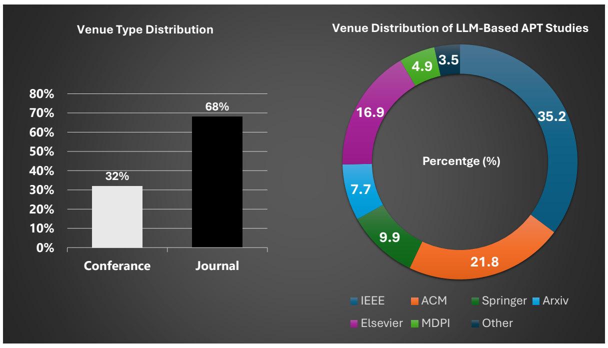

Figure 20: Distribution of Venues where LLM-based APT Studies were Published and Their Types

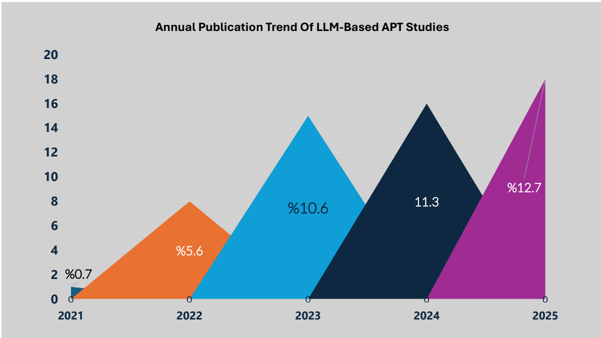

Figure 21: The Annual Trend showing the Percentage of LLM-based APT Studies published each Year

power closer to the data source, it will enable real-time data processing with low latency and low bandwidth usage [43]. However, since LLM models require high processing power and most edge devices have low capacity and low processing power, this poses a serious challenge.

Future Directions: Future researchers can work on threat-aware and edge-adaptive LLM models to overcome this challenge. For this, some techniques such as knowledge distillation, quantization, and edge-aware fine-tuning come to the fore. These techniques are explained in detail in section 5.3.

Lack of Grounded Explainability: Although LLM-based models have great potential in cybersecurity, such as APT detection, most models can cause serious security vulnerabilities in mission-critical tasks due to their black-box nature [24]. For this reason, it is necessary to understand the inputs and probabilities that the model uses when making this decision. However, there is no system that shows the necessary causal traceability and root cause reasoning information in LLM models to make this understanding. This causes gaps in forensic analysis, such as explaining attacks and auditing sources.

Future Directions: Future researchers can design more transparent systems by examining network slicing and decision-making processes for LLM models with new Explainable Artificial Intelligence (XAI) frameworks. More transparent information can be obtained with techniques that explain the training phase of models and the output phase of models, especially pre-hoc XAI and post-hoc XAI.

Scarcity of Fine-Grained LLM Training Data: Data quality has a major impact on the predictive performance of LLM models, and the volume and variety of APT-related data used in the existing literature are limited in terms of real-world representation [102]. Studies (see section 5.5) show that most studies rely on synthetic audit logs or CTIs with limited content. This limits the generalizability of LLM models and their ability to detect threats in different

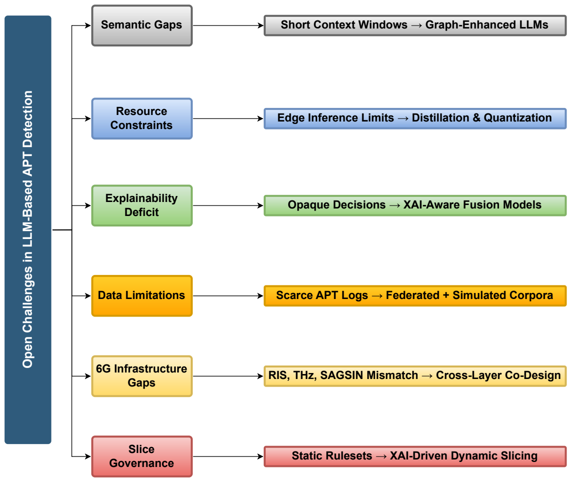

Figure 22: Taxonomy of Open Challenges in LLM-Based APT Detection

environments.

Future Directions: To overcome this data limitation issue, steps can be taken such as collaboration between researchers and organizations (public-private), development of benchmark datasets, and modeling of attack progression scenarios.

Integration with Emerging 6G Technologies: In addition to high data rates in 6G networks, new generation technologies such as intelligent reflective surfaces (IRS) and terahertz (THz) band communication will also provide more dynamic and uninterrupted communication opportunities [23]. However, this also brings some challenges such as synchronization, spectrum sharing, and secure orchestration. Adaptation of current LLM-based APT systems to such complex and multi-layered environments requires great attention.

Future Directions: To overcome these challenges, researchers can develop multi-layered security protocols. In this way, 6G networks will gain threat perception and response cycle capability in ultra-dynamic and variable environments.

Underexplored Role of Network Slicing and XAI Fusion: With the network slicing feature, 6G networks can run mission-critical scenarios such as autonomous vehicle communication and industrial control on dedicated and isolated resources [24]. However, incorrect resource allocations and unexpected load shifts that may occur during slicing operations can cause serious security problems [36]. An example is the autonomous vehicle experiencing signal delays that are beyond the delay tolerance due to network slicing.

Although XAI techniques provide dynamic adaptation capabilities in network slicing, the use of these capabilities in real-time environments is still limited [24]. Since most of the research is theory or simulation-based, its use with real-world data from SDN infrastructures needs to be investigated.

Future Directions: To overcome these limitations, researchers can develop slice-aware and network state-oriented LLM models, and these models should be able to dynamically adjust network slice configurations and allocation policies by continuously monitoring real-time data. In addition, XAI techniques can be integrated with the obtained decisions to provide traceable and reliable information for network operators.

7 CONCLUSIONS

This paper presents a comprehensive systematic review and taxonomy, the first of its kind, for LLM-based Advanced Persistent Threat (APT) detection in 6G networks. Findings from 142 recent papers examine the interaction between the capabilities (semantics) of LLM models and the challenges (architecture, privacy, etc.) of 6G environments. We aim to provide new insights for future research by presenting a taxonomy covering input types, model techniques, deployment settings, and threat lifecycle stages. Although LLM has great potential in APT attack detection, it also has limitations such as limited context memory, opaque decision processes, and real-time inference at the edge. In addition, reproducibility and dataset generalizability stand out as important obstacles for research in this area. Based on the findings, we call for joint efforts in the following research areas:

- Designing lightweight, unified LLMs for edge devices in 6G networks,
- Investigating new XAI-driven decision monitoring mechanisms to increase transparency of LLMs,
- Enriching datasets used for APT detection using fine-grained, multimodal, and real-world data,
- Integrating LLMs with slicing-aware orchestration systems in 6G for dynamic demands on 6G links.

8 Appendix A: Research Selection Criteria and Article Overview

After a comprehensive literature search and systematic analysis (Kitchenham's Systematic Literature Review (SLR) approach and Petersen's Systematic Mapping Study (SMS)), we used the form in Table 9 to select the most relevant and high-quality articles from 142 obtained articles. The questions in this form were used to select the articles focusing on LLM-based APT detection solutions among the publications addressing the Advanced Persistent Threat (APT) problem in the context of 6G wireless networks. Also, the lists of all these articles are provided in Table 10, 11, 12, 13.

Table 9: Research Evaluation Questions for LLM-Based APT Detection Studies in 6G Networks

Question	**YES**	**NO**
Does the reviewed article detect APTs in 6G networks?		
Does the reviewed article include an LLM-based solution method?		
Does the reviewed article provide an approach or methodology that includes LLMs for APT detection?		
Does the proposed method work under 6G-related constraints (such as edge resource constraints)?		
Is an optimization method suggested? If yes, which method (distillation etc.)?		
Are there datasets or simulated environments for LLM-based APT detection?		
Is the source code and/or dataset shared for reproducibility?		
What is the publication location and year?		

Acknowledgements

This work was supported by Zayed University Research Office, Fund number 23153.

References

[1] Mohamed Amine Ferrag, Othmane Friha, Burak Kantarci, Norbert Tihanyi, Lucas Cordeiro, Merouane Debbah, Djallel Hamouda, Muna Al-Hawawreh, and Kim-Kwang Raymond Choo. Edge learning for 6g-enabled internet of things: A comprehensive survey of vulnerabilities, datasets, and defenses. *IEEE Communications Surveys & Tutorials*, 25(4):2654–2713, 2023.

[2] Diana Pamela Moya Osorio, Ijaz Ahmad, José David Vega Sánchez, Andrei Gurtov, Johan Scholliers, Matti Kutila, and Pawani Porambage. Towards 6g-enabled internet of vehicles: Security and privacy. *IEEE Open Journal of the Communications Society*, 3:82–105, 2022.

[3] Sunder Ali Khowaja, Parus Khuwaja, Kapal Dev, Hussam Al Hamadi, and Engin Zeydan. Pathway to secure and trustworthy 6g for llms: Attacks, defense, and opportunities. *arXiv preprint arXiv:2408.00722*, 2024.

[4] Oscar G. Lira, Alberto Marroquin, and Marco Antonio To. Harnessing the advanced capabilities of llm for adaptive intrusion detection systems. In *International Conference on Advanced Information Networking and Applications*, pages 453–464. Springer, 2024.

[5] Burak Gulbay and Mehmet Demirci. Apt-scope: A novel framework to predict advanced persistent threat groups from enriched heterogeneous information network of cyber threat intelligence. *Engineering Science and Technology, an International Journal*, 57:101791, 2024.

[6] Md Abdur Rahman and M Shamim Hossain. A deep learning assisted software defined security architecture for 6g wireless networks: Iiot perspective. *IEEE Wireless Communications*, 29(2):52–59, 2022.

[7] Barbara Kitchenham, O Pearl Brereton, David Budgen, Mark Turner, John Bailey, and Stephen Linkman. Systematic literature reviews in software engineering–a systematic literature review. *Information and software technology*, 51(1):7–15, 2009.

[8] Kai Petersen, Robert Feldt, Shahid Mujtaba, and Michael Mattsson. Systematic mapping studies in software engineering. In *12th international conference on evaluation and assessment in software engineering (EASE)*. BCS Learning & Development, 2008.

[9] Mohammed Hassanin and Nour Moustafa. A comprehensive overview of large language models (llms) for cyber defences: Opportunities and directions. *arXiv preprint arXiv:2405.14487*, 2024.

[10] HanXiang Xu, ShenAo Wang, Ningke Li, Kailong Wang, Yanjie Zhao, Kai Chen, Ting Yu, Yang Liu, and HaoYu Wang. Large language models for cyber security: A systematic literature review. *arXiv preprint arXiv:2405.04760*, 2024.

[11] Hamza Kheddar. Transformers and large language models for efficient intrusion detection systems: A comprehensive survey. *arXiv preprint arXiv:2408.07583*, 2024.

[12] Shuang Tian, Tao Zhang, Jiqiang Liu, Jiacheng Wang, Xuangou Wu, Xiaoqiang Zhu, Ruichen Zhang, Weiting Zhang, Zhenhui Yuan, Shiwen Mao, et al. Exploring the role of large language models in cybersecurity: A systematic survey. *arXiv preprint arXiv:2504.15622*, 2025.

[13] Yiren Chen, Mengjiao Cui, Ding Wang, Yiyang Cao, Peian Yang, Bo Jiang, Zhigang Lu, and Baoxu Liu. A survey of large language models for cyber threat detection. *Computers & Security*, page 104016, 2024.

[14] Fei Zuo, Junghwan Rhee, and Yung Ryn Choe. Knowledge transfer from llms to provenance analysis: A semantic-augmented method for apt detection. *arXiv preprint arXiv:2503.18316*, 2025.

[15] Zefang Liu. A review of advancements and applications of pre-trained language models in cybersecurity. In *2024 12th International Symposium on Digital Forensics and Security (ISDFS)*, pages 1–10. IEEE, 2024.

[16] Saurabh Singh, Pradip Kumar Sharma, Seo Yeon Moon, Daesung Moon, and Jong Hyuk Park. A comprehensive study on apt attacks and countermeasures for future networks and communications: challenges and solutions. *The Journal of Supercomputing*, 75:4543–4574, 2019.

[17] Singamaneni Krishnapriya and Sukhvinder Singh. A comprehensive survey on advanced persistent threat (apt) detection techniques. *Computers, Materials & Continua*, 80(2), 2024.

[18] Zhiwei Xu, Yujuan Wu, Shiheng Wang, Jiabao Gao, Tian Qiu, Ziqi Wang, Hai Wan, and Xibin Zhao. Deep learning-based intrusion detection systems: A survey. *arXiv preprint arXiv:2504.07839*, 2025.

[19] Li-Hsiang Shen, Kai-Ten Feng, and Lajos Hanzo. Five facets of 6g: Research challenges and opportunities. *ACM Computing Surveys*, 55(11):1–39, 2023.

[20] Afroditi Blika, Stefanos Palmos, George Doukas, Vangelis Lamprou, Sotiris Pelekis, Michael Kontoulis, Christos Ntanos, and Dimitris Askounis. Federated learning for enhanced cybersecurity and trustworthiness in 5g and 6g networks: A comprehensive survey. *IEEE Open Journal of the Communications Society*, 2024.

[21] Dinh C Nguyen, Ming Ding, Pubudu N Pathirana, Aruna Seneviratne, Jun Li, Dusit Niyato, Octavia Dobre, and H Vincent Poor. 6g internet of things: A comprehensive survey. *IEEE Internet of Things Journal*, 9(1):359–383, 2021.

[22] Rawan Alghamdi, Reem Alhadrami, Dalia Alhothali, Heba Almorad, Alice Faisal, Sara Helal, Rahaf Shalabi, Rawan Asfour, Noofa Hammad, Asmaa Shams, et al. Intelligent surfaces for 6g wireless networks: A survey of optimization and performance analysis techniques. *IEEE access*, 8:202795–202818, 2020.

[23] Yang Zhao, Wenchao Zhai, Jun Zhao, Tinghao Zhang, Sumei Sun, Dusit Niyato, and Kwok-Yan Lam. A comprehensive survey of 6g wireless communications. *arXiv preprint arXiv:2101.03889*, 2020.

[24] Haochen Sun, Yifan Liu, Ahmed Al-Tahmeesschi, Avishek Nag, Mohaddeseh Soleimanpour-Moghadam, Berk Canberk, Hüseyin Arslan, and Hamed Ahmadi. Advancing 6g: Survey for explainable ai on communications and network slicing. *IEEE Open Journal of the Communications Society*, 2025.

[25] Ron Ross, Victoria Pillitteri, Gary Guissanie, Ryan Wagner, Richard Graubart, and Deborah Bodeau. Enhanced security requirements for protecting controlled unclassified information: A supplement to nist special publication 800-171. Technical report, National Institute of Standards and Technology, 2020.

[26] Kai Xing, Aiping Li, Rong Jiang, and Yan Jia. A review of apt attack detection methods and defense strategies. In *2020 IEEE Fifth International Conference on Data Science in Cyberspace (DSC)*, pages 67–70. IEEE, 2020.

[27] Diana Arulkumar, K Kartheeban, and G Arulkumaran. The apt cyber warriors with ttp weapons to battle: an review on iot and cyber twin. In *New Approaches to Data Analytics and Internet of Things Through Digital Twin*, pages 211–225. IGI Global, 2023.

[28] Yang Lv, Shaona Qin, Zifeng Zhu, Zhuocheng Yu, Shudong Li, and Weihong Han. A review of provenance graph based apt attack detection: applications and developments. In *2022 7th IEEE International Conference on Data Science in Cyberspace (DSC)*, pages 498–505. IEEE, 2022.

[29] Shamsher Ullah, Jianqiang Li, Jie Chen, Ikram Ali, Salabat Khan, Abdul Ahad, Farhan Ullah, and Victor CM Leung. A survey on emerging trends and applications of 5g and 6g to healthcare environments. *ACM Computing Surveys*, 57(4):1–36, 2024.

[30] Qi Song and Zhen Wang. An empirical study on teaching innovation of traditional wushu culture from the perspective of 6g network. *International Journal of e-Collaboration (IJeC)*, 21(1):1–18, 2025.

[31] Sri Ariyanti and Muhammad Suryanegara. Visible light communication (vlc) for 6g technology: The potency and research challenges. In *2020 Fourth world conference on smart trends in systems, security and sustainability (WorldS4)*, pages 490–493. IEEE, 2020.

[32] Mohammed Elamassie and Murat Uysal. Free space optical communication: An enabling backhaul technology for 6g non-terrestrial networks. In *Photonics*, page 1210. MDPI, 2023.

[33] Bomin Mao, Jiajia Liu, Yingying Wu, and Nei Kato. Security and privacy on 6g network edge: A survey. *IEEE communications surveys & tutorials*, 25(2):1095–1127, 2023.

[34] Pranita Bhide, Dhanush Shetty, and Suresh Mikkili. Review on 6g communication and its architecture, technologies included, challenges, security challenges and requirements, applications, with respect to ai domain. *IET Quantum Communication*, 2024.

[35] Navdeep Singh, Runumi Devi, Jay Singh, B Rajalakshmi, Mohammed Wael Mohammed, Nagarjuna Thandra, and Chalasanvi Vidyadhari. Ai-powered 6g networks: Transforming wireless communication with intelligence and terahertz waves. In *2025 International Conference on Intelligent Control, Computing and Communications (IC3)*, pages 744–750. IEEE, 2025.

[36] Ihsan H Abdulqadder and Shijie Zhou. Sliceblock: Context-aware authentication handover and secure network slicing using dag-blockchain in edge-assisted sdn/nfv-6g environment. *IEEE Internet of Things Journal*, 9(18):18079–18097, 2022.

[37] Yushan Siriwardhana, Pawani Porambage, Madhusanka Liyanage, and Mika Ylianttila. Ai and 6g security: Opportunities and challenges. In *2021 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit)*, pages 616–621. IEEE, 2021.

[38] Faisal Naeem, Mansoor Ali, Georges Kaddoum, Chongwen Huang, and Chau Yuen. Security and privacy for reconfigurable intelligent surface in 6g: A review of prospective applications and challenges. *IEEE Open Journal of the Communications Society*, 4:1196–1217, 2023.

[39] Jani Suomalainen, Ijaz Ahmad, Annette Shajan, and Tapio Savunen. Cybersecurity for tactical 6g networks: Threats, architecture, and intelligence. *Future Generation Computer Systems*, 162:107500, 2025.

[40] Christos Smiliotopoulos, Georgios Kambourakis, and Constantinos Kolias. Detecting lateral movement: A systematic survey. *Heliyon*, 10(4), 2024.

[41] Reeta Mishra, Neelu Chauduray, and Gaganjot Kaur. Harnessing machine learning for apts detection and mitigation in large-scale networks. In *2024 7th International Conference on Contemporary Computing and Informatics (IC3I)*, volume 7, pages 565–571. IEEE, 2024.

[42] Sadegh M Milajerdi, Rigel Gjomemo, Birhanu Eshete, Ramachandran Sekar, and VN Venkatakrishnan. Holmes: real-time apt detection through correlation of suspicious information flows. In *2019 IEEE symposium on security and privacy (SP)*, pages 1137–1152. IEEE, 2019.

[43] Jie Yang, Jun Zheng, Zheng Zhang, QI Chen, Duncan S Wong, and Yuanzhang Li. Security of federated learning for cloud-edge intelligence collaborative computing. *International Journal of Intelligent Systems*, 37(11):9290–9308, 2022.

[44] Astrid Bertrand, Rafik Belloum, James R Eagan, and Winston Maxwell. How cognitive biases affect xai-assisted decision-making: A systematic review. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, pages 78–91, 2022.

[45] Jie Zhang, Hui Wen, Lun Li, and Hongsong Zhu. Unittp: A unified framework for tactics, techniques, and procedures mapping in cyber threats. In *2024 IEEE 23rd International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*, pages 1580–1588. IEEE, 2024.

[46] Rishabh Agrawal, Himanshu Kumar, and Shashikant Reddy Lnu. Efficient llms for edge devices: Pruning, quantization, and distillation techniques. In *2025 International Conference on Machine Learning and Autonomous Systems (ICMLAS)*, pages 1413–1418. IEEE, 2025.

[47] Dongyuan Wu, Liming Nie, Rao Asad Mumtaz, and Kadambri Agarwal. A llm-based hybrid-transformer diagnosis system in healthcare. *IEEE Journal of Biomedical and Health Informatics*, 2024.

[48] Nikitas Karanikolas, Eirini Manga, Nikoletta Samaridi, Eleni Tousidou, and Michael Vassilakopoulos. Large language models versus natural language understanding and generation. In *Proceedings of the 27th Pan-Hellenic Conference on Progress in Computing and Informatics*, pages 278–290, 2023.

[49] Claudio Di Sipio, Riccardo Rubei, Juri Di Rocco, Davide Di Ruscio, and Ludovico Iovino. On the use of llms to support the development of domain-specific modeling languages. In *Proceedings of the ACM/IEEE 27th International Conference on Model Driven Engineering Languages and Systems*, pages 596–601, 2024.

[50] Bogdan Padiu, Radu Iacob, Traian Rebedea, and Mihai Dascalu. To what extent have llms reshaped the legal domain so far? a scoping literature review. *Information*, 15(11):662, 2024.

[51] Yiren Chen, Mengjiao Cui, Ding Wang, Yiyang Cao, Peian Yang, Bo Jiang, Zhigang Lu, and Baoxu Liu. A survey of large language models for cyber threat detection. *Computers & Security*, page 104016, 2024.

[52] Neethu Shenoy and Alex V Mbaziira. An extended review: Llm prompt engineering in cyber defense. In *2024 International Conference on Electrical, Computer and Energy Technologies (ICECET*, pages 1–6. IEEE, 2024.

[53] Wajih Ul Hassan, Adam Bates, and Daniel Marino. Tactical provenance analysis for endpoint detection and response systems. In *2020 IEEE symposium on security and privacy (SP)*, pages 1172–1189. IEEE, 2020.

[54] Yushan Liu, Xiaokui Shu, Yixin Sun, Jiyong Jang, and Prateek Mittal. Rapid: real-time alert investigation with context-aware prioritization for efficient threat discovery. In *Proceedings of the 38th Annual Computer Security Applications Conference*, pages 827–840, 2022.

[55] Robin Buchta, George Gkoktsis, Felix Heine, and Carsten Kleiner. Advanced persistent threat attack detection systems: A review of approaches, challenges, and trends. *Digital Threats: Research and Practice*, 5(4):1–37, 2024.

[56] Muhammed Golec, Sukhpal Singh Gill, Rami Bahsoon, and Omer Rana. Biosec: A biometric authentication framework for secure and private communication among edge devices in iot and industry 4.0. *IEEE Consumer Electronics Magazine*, 11(2):51–56, 2020.

[57] Muhammed Golec, Sukhpal Singh Gill, Ajith Kumar Parlikad, and Steve Uhlig. Healthfaas: Ai-based smart healthcare system for heart patients using serverless computing. *IEEE Internet of Things Journal*, 10(21):18469–18476, 2023.

[58] Houda Hafi, Bouziane Brik, Pantelis A Frangoudis, Adlen Ksentini, and Miloud Bagaa. Split federated learning for 6g enabled-networks: Requirements, challenges, and future directions. *IEEe Access*, 12:9890–9930, 2024.

[59] Meixia Tao, Yong Zhou, Yuanming Shi, Jianmin Lu, Shuguang Cui, Jianhua Lu, and Khaled B Letaief. Federated edge learning for 6g: Foundations, methodologies, and applications. *Proceedings of the IEEE*, 2024.

[60] Yifan Yao, Jinhao Duan, Kaidi Xu, Yuanfang Cai, Zhibo Sun, and Yue Zhang. A survey on large language model (llm) security and privacy: The good, the bad, and the ugly. *High-Confidence Computing*, page 100211, 2024.

[61] Xiaoliang Wang, Liming Xu, Lian Zhou, Yuzhen Liu, Neal Xiong, and Kuan-Ching Li. Large language model-driven probabilistic trajectory prediction in the internet of things using spatio-temporal encoding and normalizing flows. *Digital Communications and Networks*, 2025.

[62] Othmane Friha, Mohamed Amine Ferrag, Burak Kantarci, Burak Cakmak, Arda Ozgun, and Nassira Ghoualmi-Zine. Llm-based edge intelligence: A comprehensive survey on architectures, applications, security and trustworthiness. *IEEE Open Journal of the Communications Society*, 2024.

[63] Muhammed Golec, Guneet Kaur Walia, Mohit Kumar, Felix Cuadrado, Sukhpal Singh Gill, and Steve Uhlig. Cold start latency in serverless computing: A systematic review, taxonomy, and future directions. *ACM Computing Surveys*, 57(3):1–36, 2024.

[64] Charalampos Sergiou, Marios Lestas, Pavlos Antoniou, Christos Liaskos, and Andreas Pitsillides. Complex systems: A communication networks perspective towards 6g. *Ieee Access*, 8:89007–89030, 2020.

[65] Aditya K Sood, Sherali Zeadally, and EenKee Hong. The paradigm of hallucinations in ai-driven cybersecurity systems: Understanding taxonomy, classification outcomes, and mitigations. *Computers and Electrical Engineering*, 124:110307, 2025.

[66] Hilalah Alturkistani and Suriayati Chuprat. Artificial intelligence and large language models in advancing cyber threat intelligence: A systematic literature review. *researchsquare*, 2024.

[67] Sidahmed Benabderrahmane, Petko Valtchev, James Cheney, and Talal Rahwan. Apt-llm: Embedding-based anomaly detection of cyber advanced persistent threats using large language models. *arXiv preprint arXiv:2502.09385*, 2025.

[68] Parth Atulbhai Gandhi, Prasanna N Wudali, Yonatan Amaru, Yuval Elovici, and Asaf Shabtai. Shield: Apt detection and intelligent explanation using llm. *arXiv preprint arXiv:2502.02342*, 2025.

[69] Minghua He, Tong Jia, Chiming Duan, Huaqian Cai, Ying Li, and Gang Huang. Llmelog: An approach for anomaly detection based on llm-enriched log events. In *2024 IEEE 35th International Symposium on Software Reliability Engineering (ISSRE)*, pages 132–143. IEEE, 2024.

[70] Shiwen Shan, Yintong Huo, Yuxin Su, Yichen Li, Dan Li, and Zibin Zheng. Face it yourselves: An llm-based two-stage strategy to localize configuration errors via logs. In *Proceedings of the 33rd ACM SIGSOFT International Symposium on Software Testing and Analysis*, pages 13–25, 2024.

[71] Dincy R Arikkat, Serena Nicolazzo, Marco Arazzi, Antonino Nocera, Mauro Conti, et al. Droidttp: Mapping android applications with ttp for cyber threat intelligence. *arXiv preprint arXiv:2503.15866*, 2025.

[72] Ying Zhang, Xiaoyan Zhou, Hui Wen, Wenjia Niu, Jiqiang Liu, Haining Wang, and Qiang Li. Tactics, techniques, and procedures (ttps) in interpreted malware: A zero-shot generation with large language models. *arXiv preprint arXiv:2407.08532*, 2024.

[73] VN Ignatyev, NV Shimchik, DD Panov, and AA Mitrofanov. Large language models in source code static analysis. In *2024 Ivannikov Memorial Workshop (IVMEM)*, pages 28–35. IEEE, 2024.

[74] Hangyuan Ji, Jian Yang, Linzheng Chai, Chaoren Wei, Liqun Yang, Yunlong Duan, Yunli Wang, Tianzhen Sun, Hongcheng Guo, Tongliang Li, et al. Sevenllm: Benchmarking, eliciting, and enhancing abilities of large language models in cyber threat intelligence. *arXiv preprint arXiv:2405.03446*, 2024.

[75] Xinyu Li, Yingtong Huo, Chenxi Mao, Shiwen Shan, Yuxin Su, Dan Li, and Zibin Zheng. Anomalygen: An automated semantic log sequence generation framework with llm for anomaly detection. *arXiv preprint arXiv:2504.12250*, 2025.

[76] Junjie Huang, Zhihan Jiang, Zhuangbin Chen, and Michael R Lyu. Lunar: Unsupervised llm-based log parsing. *arXiv preprint arXiv:2406.07174*, 2024.

[77] Jule Koenders. Advancing cyberdefense through bert: A natural language processing approach for vulnerability to attack mapping within a responsible artificial intelligence framework. *Master Thesis*, 2024.

[78] Jian Wang, Tiantian Zhu, Chunlin Xiong, and Yan Chen. Multikg: Multi-source threat intelligence aggregation for high-quality knowledge graph representation of attack techniques. *arXiv preprint arXiv:2411.08359*, 2024.

[79] Dan Du, Xingmao Guan, Yuling Liu, Bo Jiang, Song Liu, Huamin Feng, and Junrong Liu. Mad-llm: A novel approach for alert-based multi-stage attack detection via llm. In *2024 IEEE International Symposium on Parallel and Distributed Processing with Applications (ISPA)*, pages 2046–2053. IEEE, 2024.

[80] Lingzhi Wang, Jiahui Wang, Kyle Jung, Kedar Thiagarajan, Emily Wei, Xiangmin Shen, Yan Chen, and Zhenyuan Li. From sands to mansions: Enabling automatic full-life-cycle cyberattack construction with llm. *arXiv preprint arXiv:2407.16928*, 2024.

[81] Shaswata Mitra, Subash Neupane, Trisha Chakraborty, Sudip Mittal, Aritran Piplai, Manas Gaur, and Shahram Rahimi. Localintel: Generating organizational threat intelligence from global and local cyber knowledge. *arXiv preprint arXiv:2401.10036*, 2024.

[82] Mouhamadou Lamine Diakhame, Chérif Diallo, and Mohamed Mejri. Mcm-llama: A fine-tuned large language model for real-time threat detection through security event correlation. In *2024 International Conference on Electrical, Computer and Energy Technologies (ICECET*, pages 1–6. IEEE, 2024.

[83] Ke Yang, Volodymyr Kindratenko, and ChengXiang Zhai. Tinyhelen's first curriculum: Training and evaluating tiny language models in a simpler language environment. *arXiv preprint arXiv:2501.00522*, 2024.

[84] Peishuai Sun, Xiaochun Yun, Shuhao Li, Tao Yin, Chengxiang Si, and Jiang Xie. Advtg: An adversarial traffic generation framework to deceive dl-based malicious traffic detection models. In *Proceedings of the ACM on Web Conference 2025*, pages 3147–3159, 2025.

[85] Jiawen Diao, Shengmin Zhao, Jianguo Xie, Rongna Xie, and Guozhen Shi. Poster: Do-hunter: A feature fusion-based llm for doh tunnel detection. In *Proceedings of the 2024 on ACM SIGSAC Conference on Computer and Communications Security*, pages 5012–5014, 2024.

[86] Hassan Jalil Hadi, Yue Cao, Sifan Li, Lexi Xu, Yulin Hu, and Mingxin Li. Real-time fusion multi-tier dnn-based collaborative idps with complementary features for secure uav-enabled 6g networks. *Expert Systems with Applications*, 252:124215, 2024.

[87] Hongbo Xu, Chengxiang Si, Chenxu Wang, Peishuai Sun, Qingyun Liu, et al. Aptsniffer: Detecting apt attack traffic using retrieval-augmented large language models. In *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5. IEEE, 2025.

[88] Aishan Liu, Yuguang Zhou, Xianglong Liu, Tianyuan Zhang, Siyuan Liang, Jiakai Wang, Yanjun Pu, Tianlin Li, Junqi Zhang, Wenbo Zhou, et al. Compromising llm driven embodied agents with contextual backdoor attacks. *IEEE Transactions on Information Forensics and Security*, 2025.

[89] Wenrui Cheng, Tiantian Zhu, Chunlin Xiong, Haofei Sun, Zijun Wang, Shunan Jing, Mingqi Lv, and Yan Chen. Sok: Knowledge is all you need: Last mile delivery for automated provenance-based intrusion detection with llms. *arXiv preprint arXiv:2503.03108*, 2025.

[90] Weilan Wang, Yu Mao, Dongdong Tang, Hongchao Du, Nan Guan, and Chun Jason Xue. When compression meets model compression: Memory-efficient double compression for large language models. *arXiv preprint arXiv:2502.15443*, 2025.

[91] Wentao Zhao, Wenpeng Jing, Zhaoming Lu, and Xiangming Wen. Edge and terminal cooperation enabled llm deployment optimization in wireless network. In *2024 IEEE/CIC International Conference on Communications in China (ICCC Workshops)*, pages 220–225. IEEE, 2024.

[92] Zhichao Xu, Ashim Gupta, Tao Li, Oliver Bentham, and Vivek Srikumar. Beyond perplexity: Multi-dimensional safety evaluation of llm compression. *arXiv preprint arXiv:2407.04965*, 2024.

[93] Wenxiao Wang, Wei Chen, Yicong Luo, Yongliu Long, Zhengkai Lin, Liye Zhang, Binbin Lin, Deng Cai, and Xiaofei He. Model compression and efficient inference for large language models: A survey. *arXiv preprint arXiv:2402.09748*, 2024.

[94] Dong Liu, Yanxuan Yu, Yite Wang, Jing Wu, Zhongwei Wan, Sina Alinejad, Benjamin Lengerich, and Ying Nian Wu. Designing large foundation models for efficient training and inference: A survey. *arXiv preprint arXiv:2409.01990*, 2024. Version 5, revised 14 Apr 2025.

[95] Ruiyang Qin, Dancheng Liu, Chenhui Xu, Zheyu Yan, Zhaoxuan Tan, Zhenge Jia, Amir Nassereldine, Jiajie Li, Meng Jiang, Ahmed Abbasi, et al. Empirical guidelines for deploying llms onto resource-constrained edge devices. *arXiv preprint arXiv:2406.03777*, 2024.

[96] Chang Liu and Jun Zhao. Resource allocation for stable llm training in mobile edge computing. In *Proceedings of the Twenty-fifth International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing*, pages 81–90, 2024.

[97] Zheming Yang, Yuanhao Yang, Chang Zhao, Qi Guo, Wenkai He, and Wen Ji. Perllm: Personalized inference scheduling with edge-cloud collaboration for diverse llm services. *arXiv preprint arXiv:2405.14636*, 2024.

[98] Benedetta Picano, Dinh Thai Hoang, and Diep Nguyen. A matching game for llm layer deployment in heterogeneous edge networks. *IEEE Open Journal of the Communications Society*, 2025.

[99] Guanqiao Qu, Qiyuan Chen, Wei Wei, Zheng Lin, Xianhao Chen, and Kaibin Huang. Mobile edge intelligence for large language models: A contemporary survey. *IEEE Communications Surveys & Tutorials*, 2025.

[100] Mingjin Zhang, Xiaoming Shen, Jiannong Cao, Zeyang Cui, and Shan Jiang. Edgeshard: Efficient llm inference via collaborative edge computing. *IEEE Internet of Things Journal*, 2024.

[101] Sowmya Myneni, Kritshekhar Jha, Abdulhakim Sabur, Garima Agrawal, Yuli Deng, Ankur Chowdhary, and Dijiang Huang. Unraveled—a semi-synthetic dataset for advanced persistent threats. *Computer Networks*, 227:109688, 2023.

[102] Erfan Ghiasvand, Suprio Ray, Shahrear Iqbal, Sajjad Dadkhah, and Ali A Ghorbani. Cicapt-iiot: A provenance-based apt attack dataset for iiot environment. *arXiv preprint arXiv:2407.11278*, 2024.

[103] Yi-Ting Huang, Ying-Ren Guo, Yu-Sheng Yang, Guo-Wei Wong, Yu-Zih Jheng, Yeali Sun, Jessemyn Modini, Timothy Lynar, and Meng Chang Chen. Saga: Synthetic audit log generation for apt campaigns. *arXiv preprint arXiv:2411.13138*, 2024.

[104] Samaneh Shafee, Alysson Bessani, and Pedro M Ferreira. Evaluation of llm-based chatbots for osint-based cyber threat awareness. *Expert Systems with Applications*, page 125509, 2024.

[105] Abdullah Said AL-Aamri, Rawad Abdulghafor, Sherzod Turaev, Imad Al-Shaikhli, Akram Zeki, and Shuhaili Talib. Machine learning for apt detection. *Sustainability*, 15(18):13820, 2023.

[106] Helmut Neuschmied, Martin Winter, Branka Stojanović, Katharina Hofer-Schmitz, Josip Božić, and Ulrike Kleb. Apt-attack detection based on multi-stage autoencoders. *Applied Sciences*, 12(13):6816, 2022.

[107] Haider W Oleiwi, Doaa N Mhawi, and Hamed Al-Raweshidy. A meta-model to predict and detect malicious activities in 6g-structured wireless communication networks. *Electronics*, 12(3):643, 2023.

[108] Mamoon M Saeed, Rashid A Saeed, Maha Abdelhaq, Raed Alsaqour, Mohammad Kamrul Hasan, and Rania A Mokhtar. Anomaly detection in 6g networks using machine learning methods. *Electronics*, 12(15):3300, 2023.

[109] Cho Do Xuan and Tung Thanh Nguyen. A novel approach for apt attack detection based on an advanced computing. *Scientific Reports*, 14(1):22223, 2024.

[110] Hoa Cuong Nguyen, Cho Do Xuan, Long Thanh Nguyen, and Hoa Dinh Nguyen. A new framework for apt attack detection based on network traffic. *Journal of Intelligent & Fuzzy Systems*, 44(3):3459–3474, 2023.

[111] Chunlin Xiong, Tiantian Zhu, Weihao Dong, Linqi Ruan, Runqing Yang, Yueqiang Cheng, Yan Chen, Shuai Cheng, and Xutong Chen. Conan: A practical real-time apt detection system with high accuracy and efficiency. *IEEE Transactions on Dependable and Secure Computing*, 19(1):551–565, 2020.

[112] Muhammad Mudassar Yamin, Ehtesham Hashmi, Mohib Ullah, and Basel Katt. Applications of llms for generating cyber security exercise scenarios. *IEEE Access*, 2024.

Table 10: Overview of 58 Studies Related to RQ1

No	Paper	Code Availability	Dataset	Eval. Protocol	Venue/Platform	Year
1	Koenders (1.pdf)	NO	CVE/ATT&CK	YES	Erasmus MSc	2024
2	Zuo et al. (2.pdf)	NO	audit data implied	YES	Sandia/UCO	2025
3	Cheng et al. (3.pdf)	GitHub	benchmark	OMNISEC eval	arXiv	2025
4	Xu et al. (4.pdf)	NO	public datasets	survey w/ metrics	ACM JACM	2025
5	Ali et al. (5.pdf)	NO	NO	NO	SHIFRA Journal	2025
6	Zhang et al. (6.pdf)	NO	NO	NO	TechRxiv (preprint)	2025
7	Jeon et al. (7.pdf)	NO	NO	RAG-based graph eval	Conf. Paper (Korea)	2025
8	Blänsdorf (8.pdf)	NO	CTI list	manual labeling	MSc Thesis (Chalmers)	2024
9	Yin et al. (9.pdf)	NO	NO	NO	arXiv	2025
10	Antar (10.pdf)	ECHO env.	NO	PromptPilot eval	MSc Thesis (Queen's)	2025
11	Xu et al. (11.pdf)	NO	CTI reports	IntelEX framework	arXiv	2025
12	Chen et al. (12.pdf)	AECR pipeline	NO	F1, precision, recall	Elsevier CoSE	2025
13	Tan et al. (13.pdf)	NO	NO	temporal graph eval	Conf. (Glasgow)	2025
14	Tan et al. (14.pdf / 16.pdf)	NO	NO	survey + taxonomy	IEEE IoT Journal	2025
15	Purba (15.pdf)	NO	NO	Kibana-query generation	PhD Diss. (UNC Charlotte)	2025
16	Wang et al. (17.pdf)	pending	APT dataset ¿1k	AURORA system	arXiv	2025
17	Kavousi (18.pdf)	NO	NO	semantic security eval	PhD Diss. (Northwestern)	2025
18	Zhang & Tenney (19.pdf)	NO	NO	NO (survey only)	OJBM	2024
19	Daniel et al. (20.pdf)	NO	Snort rules	LLM vs. ML eval	MDPI BDCC	2025
20	Ahmed (21.pdf)	prototype impl.	DARPA OpTC	distributed eval	PhD Diss. (UNC Charlotte)	2024
21	Du et al. (22.pdf)	NO	benchmark used	MAD-LLM eval	IEEE ISPA	2024
22	Mezzi et al. (23.pdf)	Eval framework	350 CTI reports	calibration, consistency	arXiv	2025
23	Suomalainen et al. (24.pdf)	NO	Cyber Ops Tracker	LLM for CTI metrics	TechRxiv	2025
24	Alturkistani (25.pdf)	NO	NO	systematic SLR analysis	Research Square (SLR)	2024
25	Sultana et al. (26.pdf)	NO	NO	LLM eval. framework	IEEE CNS Workshop	2023
26	Cui et al. (27.pdf)	NO	NO	LLM risk tax. + eval	Tsinghua Lab + Ant Group	2024
27	Li et al. (28.pdf)	NO	NO	agentic eval + 5G	NYU Tech Report (arXiv)	2025
28	Daniel et al. (29.pdf)	Snort parser	973 rules dataset	LLM vs ML accuracy	arXiv	2024
29	Mitra et al. (30.pdf)	NO	NO	LocalIntel eval	arXiv	2025
30	Wang et al. (31.pdf) – MultiKG	MultiKG repo	real CTI + logs	cross-source KG eval	arXiv	2024
31	Yao (32.pdf)	5GSecRec impl.	5G+Kube alerts	QA + correlation	MSc Thesis (Concordia)	2024
32	Mahboubi et al. (33.pdf)	NO	open ontologies	survey + ML eval	Elsevier JNCA	2024
33	Sammouri (34.pdf) – CAPEC model	CAPEC model	CAPEC taxonomy	expert evaluation	MSc Thesis (Miami Univ.)	2025
34	Sewak et al. (35.pdf)	NO	NO	threat graph eval	CIKM Workshop	2023
35	Kasri et al. (36.pdf)	NO	NO	review + cases	MDPI Computation	2025
36	Hasanov et al. (37.pdf)	NO	NO	SLR criteria	IEEE Access	2024
37	Keltek et al. (38.pdf) – LSAST	LSAST prototype	HackerOne dump	LLM vs SAST	arXiv	2024
38	Jawad (39.pdf)	NO	NO	PhD eval phases	PhD Plan (Spain)	2025
39	Würsch et al. (40.pdf)	NO	arXiv NLP corpus	NER/ER comparison	arXiv	2023
40	Diakhame et al. (41.pdf) – MCM-Llama	LLM pipeline	security events	LLM vs NER/Sim	ICECET (IEEE Conf.)	2024
41	Chen et al. (42.pdf) – LLM Survey	NO	NO	systematic stages review	Elsevier CoSE	2024
42	Rahman (43.pdf) – Incident Reconstruct	local LLM sys.	PCAP, ChromaDB	LLM eval on NTA	MSc Thesis (Turku)	2024
43	Ji et al. (44.pdf) – SEVENLLM	GitHub	28-task benchmark	multi-task eval	arXiv	2024
44	Taghavi et al. (45.pdf) – LLM4Vuln	NO	NO	survey + workflow eval	ResearchSquare (preprint)	2024
45	Sood et al. (46.pdf) – Hallucination	NO	NO	taxonomy + mitigation	Elsevier CEE	2025
46	Zhang et al. (47.pdf) – GENTTP	Tool released	PyPI Malware + GT	Zero-shot eval, chatbot	arXiv	2024
47	Fayyazi et al. (48.pdf) – TTP-LLM	GitHub	NO	RAG vs SFT benchmark	IEEE ACSAC Workshops	2024
48	Zhang et al. (49.pdf) – UniTTP	NO	Internal + public	F1, task-level	IEEE TrustCom	2024
49	Arikkata et al. (50.pdf) – DroidTTP	RAG pipeline	Android TTP Dataset	Jaccard, Hamming Loss	arXiv	2025
50	Li et al. (51.pdf) – AnomalyGen	LogSynth code	synthetic logs	F1 gain eval	Conf. Paper (preprint)	2025
51	Shan et al. (52.pdf) – LogConfigLocal	tool available	Hadoop logs	root-cause accuracy	ACM ISSTA	2024
52	Huang et al. (53.pdf) – LUNAR	GitHub	LCU logs	log parsing eval	ASE Conf. (arXiv preprint)	2024
53	He et al. (54.pdf) – LLMeLog	fine-tuned BERT	3 public logs	F1 ¿ 99%	IEEE ISSRE	2024
54	Wang et al. (55.pdf) – LM Agents	NO	NO	multi-agent eval	arXiv	2025
55	Balasubramanian et al. (56.pdf)	chatbot code	open logs used	GPT-3 vs others eval	IEEE BigData	2023
56	Gandhi et al. (57.pdf) – SHIELD	NO	NO (custom logs only)	precision/recall	arXiv	2025
57	Benabderrahmane et al. (58.pdf)	NO	DARPA TC	AE/VAE eval	arXiv	2025
58	Ferrag et al. (59.pdf) – LLM Survey	NO	NO	survey w/ benchmarks	SSRN (preprint)	2025

Table 11: Overview of 28 Studies Related to RQ2

No	Paper	Code Availability	Dataset	Eval. Protocol	Venue/Platform	Year
1	Albshaier et al. (Ä.pdf)	NO	Not specified (SLR only)	YES	Electronics (MDPI)	2025
2	Hameed et al. (A.pdf)	NO	IoT-based (SMPC, DP, HE)	Preprint (SSRN)	SSRN / Nuclear Phys. B	2025
3	Al-Kadhimi et al. (a1.pdf)	NO	Mobile APT data (1351 reviewed)	YES (SLR + Framework)	Applied Sciences (MDPI)	2023
4	Le & Shetty (B.pdf)	NO	5G-based IoT	Conceptual (no benchmarks)	Ad Hoc Networks (Elsevier)	2021
5	Alkaeed et al. (C.pdf)	NO	AI-XR / Metaverse data	YES (Survey)	J. of Network and Comp. Apps	2024
6	Xu et al. (Ç.pdf)	NO	Multiple (127 reviewed)	YES (Survey)	LLM4Security Survey	2024
7	Zhang et al. (D.pdf)	YES (planned)	4 RS models (LLM-based)	Experiments on 4 models	arXiv	2024
8	Guo et al. (E.pdf)	YES	Various LLMs (LLaMA, GPT)	Extensive experiments	ICML (PMLR)	2024
9	Cao et al. (F.pdf)	NO	Fine-tuned LLMs (backdoor)	Experiments (safety test bypass)	arXiv	2024
10	Aguilera-Martínez & Berzal (G.pdf)	NO	Training + Inference threats	YES (Survey)	arXiv	2025
11	Lanka et al. (O.pdf)	NO	Honeypot + UEBA data	YES (LLM-based analysis)	Electronics (MDPI)	2024
12	Zhang et al. (H.pdf)	NO	Poisoned RAG documents	YES	FSE Companion (ACM)	2024
13	Rahman & Hossain (I.pdf)	NO	IIoT logs in 6G	YES (SDS + DL)	IEEE Wireless Comm.	2022
14	Wang et al. (ı.pdf)	NO	MEC + AI security logs	YES (ETSI-based survey)	IEEE IoT Journal	2023
15	Alevizos et al. (İ.pdf)	NO	Blockchain-based IDS in VSNs	YES (Throttling eval)	Sensors (MDPI)	2023
16	Nahar et al. (J.pdf)	NO	ZTA in 6G	YES (Use case studies)	IEEE Access	2024
17	Je et al. (K.pdf)	NO	Open 6G + AI systems	YES (Threat mapping)	IEEE Comm. Standards Mag.	2021
18	Xu et al. (L.pdf)	YES	APT Traffic (Anyrun2024)	YES	CAS / UCAS	2024
19	Du et al. (M.pdf)	NO	Multi-source alerts	YES	IEEE ISPA 2024	2024
20	Liu et al. (N.pdf)	YES	Contextual demos for agents	YES (Backdoor trigger eval)	IEEE TIFS	2025
21	Hassanin & Moustafa (Z.pdf)	NO	Survey on LLM for cyber defense	YES	arXiv	2024
22	Mao et al. (Ö.pdf)	NO	Edge computing / cache / intelligence	YES	IEEE COMST	2023
23	Yang et al. (P.pdf)	NO	6G Security Protocols	YES	arXiv / ACM	2024
24	Hadi et al. (R.pdf)	YES	UAVIDS / NF-UQ / 5G-NIDD	YES	Expert Systems w/ Applications	2024
25	Chen et al. (Ş.pdf)	NO	Various threat logs + LLMs	YES	Computers & Security	2024
26	Sun et al. (T.pdf)	YES	DL model evasions (semantic traffic)	YES	WWW 2025	2025
27	Diao et al. (V.pdf)	NO	DoH tunnel traffic	YES (Feature fusion, Recall: 0.9995)	ACM CCS Poster	2024
28	Sun et al. (Y.pdf)	YES	Adversarial LLM traffic (6 datasets)	YES (RL + Payload tuning)	WWW 2025	2025

Table 12: Overview of 26 Studies Related to RQ3

No	Paper	Code Availability	Dataset	Eval. Protocol	Venue/Platform	Year
1	Liu et al. (a1.pdf)	NO	Not specified	Review (model compression only)	Frontiers in Robotics and AI	2025
2	Friha et al. (a2.pdf)	NO	Not specified	Comprehensive survey	IEEE OJCOMS	2024
3	Zhang et al. (a3.pdf)	NO	Llama2	Real testbed with Llama2 models	IEEE IoT Journal	2025
4	Cai et al. (a4.pdf)	NO	LLaMA, ChatGLM	Edge-LLM framework eval	Conference (not stated)	2025
5	Qu et al. (a10.pdf)	NO	Not specified	Survey on MEI for LLMs	IEEE COMST (accepted)	2025
6	Picano et al. (a6.pdf)	NO	Testbed for autonomous driving	Matching-based optimization	IEEE OJCOMS	2025
7	Ray & Pradhan (a9.pdf)	NO	IoT edge, quantized LLMs	LLMEdge Framework Demo	Not stated	2025
8	Kim et al. (a20.pdf)	NO	Not specified	Systematic review on compression & tuning	ACM Computing Surveys	2025
9	Zhang et al. (a11.pdf)	NO	Not specified	Quantization + batching (wireless constraint)	IEEE TWC	2025
10	Wei et al. (a12.pdf)	YES (T-MAC GitHub)	LLaMA, BitNet	LUT-based low-bit benchmarking	EuroSys 2025	2025
11	Semerikov et al. (a13.pdf)	NO	LLM-edge cases, EdgeLLM, EdgeShard	Comprehensive Survey	CEUR Workshop Proceedings	2025
12	Dhar et al. (a14.pdf)	NO	LLaMA-2 7B INT4	Empirical edge inference eval	ACMSE 2024	2024
13	Zheng et al. (a15.pdf)	NO	Meta LLaMA, DeepSeek	Lifecycle review + hardware co-design	ACM Computing Surveys	2025
14	Qin et al. (a16.pdf)	NO	LaMP datasets	Empirical guidelines + compression comparison	arXiv (Preprint)	2025
15	Yang et al. (a17.pdf)	NO	Not specified	PerLLM edge-cloud scheduling	arXiv (Preprint)	2024
16	Liu et al. (a18.pdf)	NO	Not specified	Fractional programming optimization	MOBIHOC '24	2024
17	Zhu et al. (a19.pdf)	NO	Not specified	Compression taxonomy (quant., prune, KD)	TACL	2024
18	Zhao et al. (a29.pdf)	NO	Not specified	Edge-terminal token decoding optimization	IEEE (Wireless)	2024
19	Wang et al. (a21.pdf)	NO	Not specified	Inference taxonomy (compression focus)	IEEE (Preprint)	2024
20	Wang et al. (a22.pdf)	NO	Quantized OPT-1.3B	Compression-aware quantization + pruning	arXiv (Preprint)	2025
21	Xu et al. (a23.pdf)	YES (GitHub)	Multiple LLMs	Multi-dimensional safety evaluation	arXiv (Preprint)	2024
22	Liu et al. (a24.pdf)	NO	GPTQ, SmoothQuant	Survey on efficient training/inference	arXiv (Preprint)	2025
23	Wen et al. (a25.pdf)	NO	Custom telemetry logs	LLM-based anomaly detection in 6G	HOTNETS '24	2024
24	Khowaja et al. (a26.pdf)	NO	Named Entity Recognition (NER)	Membership inference on ZSM fine-tuning	IEEE (Preprint)	2024
25	Qin et al. (a27.pdf)	NO	Multiple (SAGIN datasets)	CoT-based security for 6G SAGIN	IEEE (Preprint)	2025
26	Qu et al. (a28.pdf)	NO	MEI4LLM (LLMs + Edge)	Survey + MEI framework	IEEE COMST (Accepted)	2025

Table 13: Overview of 30 Studies Related to RQ4

No	Paper	Code Availability	Dataset	Eval. Protocol	Venue/Platform	Year
1	Abu Talib et al. (2.pdf)	NO	YES	YES (Systematic Review)	Computers & Security	2022
2	Abu Talib et al. (23.pdf)	NO	NO	YES (Review on APT Beaconing)	Computers & Security	2022
3	Al-Aamri et al. (32.pdf)	NO	YES (Custom logs + flow)	YES (CDT model + IDS)	Sustainability (MDPI)	2023
4	Do Xuan & Nam (3.pdf)	NO	NO	YES (Domain monitoring)	Procedia Computer Science	2019
5	Do Xuan & Nguyen (30.pdf)	NO	YES (APT IP traffic)	YES (BiLSTM + Attention + DGCNN)	Scientific Reports	2024
6	Do Xuan et al. (4.pdf)	NO	YES (Reconstructed flows)	YES (BiLSTM-GCN based)	J. Intelligent & Fuzzy Syst.	2020
7	El Alami & Rawat (16.pdf)	NO	YES (TON-IoT)	YES (GAN, LSTM, AE eval)	—	2024
8	Ferrag et al. (11.pdf)	NO	YES (FL/Edge datasets)	YES (Comprehensive Survey)	IEEE Com. Surveys & Tutorials	2023
9	Ferrag et al. (27.pdf)	NO	YES (42 models & datasets)	YES (Taxonomy + LLM Eval)	SSRN (preprint)	2024
10	Ghiasvand et al. (25.pdf)	YES	YES (CICAPT-IIoT Dataset)	YES (Multi-phase APT dataset)	arXiv	2024
11	Gupta et al. (15.pdf)	YES (PyTorch model)	YES (KDDCup)	YES (DoS, Probe, Sybil attacks)	—	2024
12	Huang et al. (24.pdf)	YES (SAGA)	YES (Synthetic Logs)	YES (Technique-hunting, APT lifecycle)	—	2024
13	Motlagh et al. (19.pdf)	NO	NO	YES (Offensive & defensive use)	arXiv	2024
14	Myneni et al. (20.pdf)	NO	YES (Unraveled dataset)	YES (Semi-synthetic APT eval)	Computer Networks	2023
15	Neuschmied et al. (31.pdf)	NO	YES (Contagio + CICIDS2017)	YES (AE + VAE models)	Applied Sciences (MDPI)	2022
16	Nezhadsistani & Stiller (8.pdf)	NO	YES	YES (Survey, Challenges, Metrics)	IEEE 6GNet	2024
17	Nguyen et al. (17.pdf)	NO	NO	YES (Threat taxonomy, LLMSecOps)	arXiv	2024
18	Nguyen et al. (5.pdf)	NO	YES	YES (MIG: MLP + Inference + GCN)	J. Intelligent & Fuzzy Syst.	2023
19	Oleiwi et al. (28.pdf)	NO	YES (NSL-KDD, CIC, etc.)	YES (Stacked meta-model, classifiers)	Electronics (MDPI)	2023
20	Rajendran & Vyas (6.pdf)	YES	YES (Custom)	YES (Comparative Eval)	SoutheastCon	2024
21	Saeed et al. (29.pdf)	NO	YES (Multiple APT datasets)	YES (Hybrid EL, CFS-RF, Adaboost)	Electronics (MDPI)	2023
22	Shafee et al. (18.pdf)	NO	YES (Twitter-CTI)	YES (Binary Class. + NER)	Expert Syst. with Applications	2025
23	Sharma & Rani (14.pdf)	NO	YES (RT-IoT)	YES (Stacked-Hybrid ML eval)	IEEE IoT Journal	2024
24	Stojanović et al. (1.pdf)	NO	YES	YES (Review of 20+ datasets)	Computers & Security	2020
25	Unnamed (22.pdf)	NO	YES (Semi-synthetic network traffic)	YES (ML early detection comparison)	LLM-Aided (Unspecified)	2024
26	Viswanathan et al. (21.pdf)	NO	YES (Partially synthetic MRI)	YES (ML vs full/partial synthetic eval)	BioMed (MRI Imaging)	2024
27	Xiong et al. (13.pdf)	NO	YES (Windows-host logs)	YES (FSM + Real-time eval)	IEEE TDSC	2022
28	Xylouris et al. (7.pdf)	YES (XGBoost, CNN, etc.)	YES (5G Testbed Dataset)	YES (Real-time SHAP Eval)	IEEE TCE	2024
29	Yamin et al. (10.pdf)	YES (CyExec - GPT)	NO	YES (Scenario Gen, RAG prompts)	IEEE Access	2024
30	Zhang et al. (26.pdf)	NO	YES	YES (PADASYN + AdaBoost eval)	—	2024