Large Language Models for Computer Networking Operations and Management: A Survey on Applications, Key Techniques, and Opportunities

Fan Liu[1], Behrooz Farkiani[1], and Patrick Crowley[1]

[1]Washington University in St. Louis St. Louis

June 06, 2025

Abstract

This survey examines the application of Large Language Models (LLMs) in network operations and management (NO&M). It outlines the transformation in NO&M driven by LLMs, highlighting their potential to address challenges across network design, automation, optimization, and security domains. The paper explores how LLMs enhance traditional methods by automating complex tasks, improving network agility, and providing solutions to emerging network demands. We present our methodology for a systematic literature review and analyze how LLMs complement network technologies including Software-Defined Networking, Network Function Virtualization, Intent-Based Networking, and Zero-Touch Networks. The survey categorizes existing research into key application areas, providing comparisons between LLM-based approaches and traditional methods. We identify current limitations, such as integration with legacy systems, explainability, data privacy, and computational scalability. Additionally, we propose future research directions, including domain-specific efficient architectures, advanced intent-based management, privacy-preserving techniques, integration with next-generation networks, sustainable LLM solutions, cross-domain collaboration frameworks, and ethical considerations. Our findings offer insights for researchers and practitioners aiming to leverage LLMs for intelligent network management in complex and dynamic environments.

Large Language Models for Computer Networking Operations and Management: A Survey on Applications, Key Techniques, and Opportunities

Fan Liu
Washington University in St. Louis
St. Louis, Missouri, USA
fan.liu@wustl.edu

Behrooz Farkiani
Washington University in St. Louis
St. Louis, Missouri, USA
b.farkiani@wustl.edu

Patrick Crowley
Washington University in St. Louis
St. Louis, Missouri, USA
pcrowley@wustl.edu

Abstract

This survey examines the application of Large Language Models (LLMs) in network operations and management (NO&M). It outlines the transformation in NO&M driven by LLMs, highlighting their potential to address challenges across network design, automation, optimization, and security domains. The paper explores how LLMs enhance traditional methods by automating complex tasks, improving network agility, and providing solutions to emerging network demands. We present our methodology for a systematic literature review and analyze how LLMs complement network technologies including Software-Defined Networking, Network Function Virtualization, Intent-Based Networking, and Zero-Touch Networks. The survey categorizes existing research into key application areas, providing comparisons between LLM-based approaches and traditional methods. We identify current limitations, such as integration with legacy systems, explainability, data privacy, and computational scalability. Additionally, we propose future research directions, including domain-specific efficient architectures, advanced intent-based management, privacy-preserving techniques, integration with next-generation networks, sustainable LLM solutions, cross-domain collaboration frameworks, and ethical considerations. Our findings offer insights for researchers and practitioners aiming to leverage LLMs for intelligent network management in complex and dynamic environments.

Keywords

Large Language Models, Computer Networking, Generative AI, Network Intelligence, Network Management, Network Operations

NOMENCLATURE

5G Fifth Generation
6G Sixth Generation
ABR Adaptive Bitrate
AI Artificial Intelligence
APT Advanced Persistent Threat
BERT Bidirectional Encoder Representations from Transformers
BLEU Bilingual Evaluation Understudy
CLI Command-Line Interface
DDoS Distributed Denial of Service
DL Deep Learning
DRL Deep Reinforcement Learning
EGI Edge General Intelligence
FCAPS Fault, Configuration, Accounting, Performance, Security
FQDN Fully Qualified Domain Name
FSM Finite State Machine
GAI Generative AI
GAT Graph Attention Network
GMM Gaussian Mixture Models
GNN Graph Neural Network
GPT Generative Pre-trained Transformer
IBN Intent-Based Networking
IDS Intrusion Detection System
IoT Internet of Things
IRAG Intent-based Retrieval Augmented Generation
ISTN Integrated Satellite-Terrestrial Network
KNN K-Nearest Neighbors
LLM Large Language Model
LoRA Low-Rank Adaptation
LSTM Long Short-Term Memory
ML Machine Learning
MLP Multi-Layer Perceptron
MoE Mixture of Experts
NDT Network Digital Twin
NF Network Function
NFV Network Function Virtualization
NIDS Network Intrusion Detection System
NLP Natural Language Processing
NO&M Network Operations and Management
NSD Network Service Descriptor
NSGA2 Non-dominated Sorting Genetic Algorithm II
PPDIOO Prepare, Plan, Design, Implement, Operate, Optimize
QoE Quality of Experience
QoS Quality of Service
RAG Retrieval-Augmented Generation
RFC Request for Comments
SAGIN Space-Air-Ground Integrated Network
SDN Software-Defined Networking
SFC Service Function Chain
SLA Service Level Agreement
SVM Support Vector Machine
UAV Unmanned Aerial Vehicle
VLAN Virtual Local Area Network
XAI Explainable AI
ZTN Zero-Touch Network

1 Introduction

The advancement of network technologies has revolutionized how we communicate, share information, and stay connected in an increasingly interconnected world [53, 116, 119]. As modern infrastructures grow in complexity, NO&M faces increasing demands for

scalability, adaptability, and automation. Traditional network management approaches, which rely heavily on manual configurations and rule-based automation, struggle to keep up with the pace of evolving network environments. These methods are error-prone, time-consuming, and difficult to scale, making them inadequate for large-scale or dynamic infrastructures. Furthermore, they lack real-time adaptability, necessitating human oversight and slowing down decision-making [69].

To address these challenges, AI has been successfully integrated into NO&M, particularly with ML techniques that automate network operations, optimize performance, and enhance security [111, 136]. AI-driven models have been widely deployed in 5G networks, improving traffic optimization, anomaly detection, and resource allocation [18, 49]. However, while these techniques excel at domain-specific tasks, they often require extensive task-specific training, frequent fine-tuning, and large labeled datasets to adapt to changing network conditions [122]. Some AI models also struggle with unstructured network data, leading to integration challenges within existing network infrastructures.

The emergence of LLMs presents a new approach to addressing some of these limitations. Unlike traditional AI models, LLMs can generalize across multiple tasks, and interact with human operators using natural language [141, 151]. This capability introduces a more intuitive way to manage and automate networks, allowing engineers to issue commands in natural language rather than writing complex scripts. For instance, an LLM-powered assistant could answer "Why is loss spiking on port Gi1/0/24?" with "The interface is experiencing a high packet drop rate due to excessive buffer utilization—consider adjusting the queueing strategy or increasing buffer size." This shift enhances the efficiency of network troubleshooting, reducing manual errors and improving operational decision-making.

Furthermore, techniques such as prompt engineering, few-shot learning, and RAG allow users to adapt LLMs to new tasks without requiring full retraining of the models [82, 86, 138, 142]. While LLMs demand significant computational resources for training, their ability to reduce the frequency of retraining in operational settings can lower long-term adaptation costs compared to traditional ML models [10]. However, LLM inference costs are generally higher than traditional ML inference costs due to the LLMs' massive size and complexity, requiring more compute resources and specialized hardware like GPUs or TPUs, leading to increased operational expenses [81].

LLMs are not meant to replace traditional AI models but rather to complement them by offering stronger natural language capabilities, reduced need for retraining and adaptation to different domains.

This survey investigates the role of LLMs in NO&M across four key domains: Network Design, Network Automation, Network Optimization, and Network Security. We investigated a total of 67 papers following our methodology described in Section 2. As shown in Figure 1, the number of research publications in this field has grown rapidly, from just 1 paper in 2021 to 42 papers in 2024, highlighting the increasing interest in LLM-driven networking solutions. Although our data may not represent the exact number of all published papers in this field, the trend is clearly visible from our year-by-year distribution analysis. By providing an analysis of their applications, limitations, and future potential, this survey aims to bridge the gap between traditional AI-driven NO&M and the next generation of intelligent, autonomous networks.

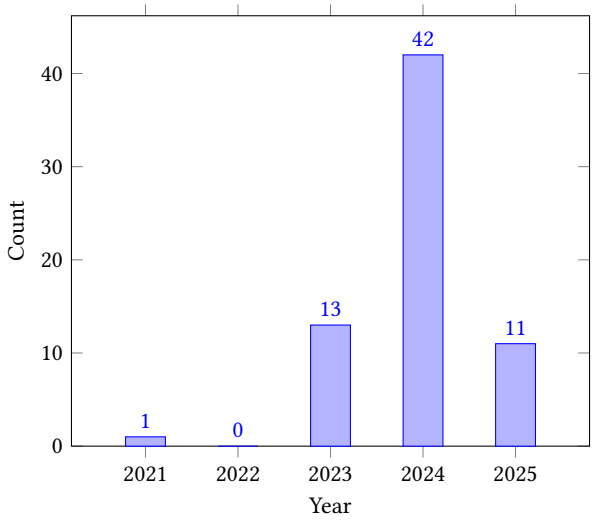

Figure 1: Number of publications exploring LLMs in networking from 2021 to 2025 (based on the papers investigated in this study).

These four domains represent concrete, practical activities within the broader NO&M lifecycle. While general frameworks like FCAPS (Fault, Configuration, Accounting, Performance, and Security) [145] or Cisco's PPDIOO (Prepare, Plan, Design, Implement, Operate, and Optimize) [28] provide models for Network Management, our categorization focuses on essential operational areas where LLMs are making impacts, aligning with current networking challenges and research directions."

Our contributions can be summarized as follows:

- **Systematic Analysis of LLM Applications in NO&M** This survey provides a structured analysis of how LLMs are currently applied across key network domains, demonstrating their practical impacts on network design, automation, optimization, and security. We include quantitative comparisons between LLM-based approaches and traditional methods to illustrate performance differences.
- **Integration Framework for LLMs with Network Technologies:** We present a structured analysis of how LLMs complement and enhance modern networking paradigms including SDN, NFV, IBN, and ZTN. This framework helps practitioners understand the varying levels of integration maturity across different technologies and identify promising opportunities.
- **Technical Challenges and Practical Limitations:** The survey identifies key implementation challenges such as computational scalability, integration with legacy systems, and domain knowledge acquisition. We provide concrete examples from current research demonstrating how these limitations manifest in real-world networking applications.
- **Research Roadmap with Actionable Directions:** Based on our analysis of current limitations, we propose a detailed

research roadmap emphasizing targeted improvements in areas including domain-specific architectures, intent-based management, and cross-domain collaboration frameworks. Each direction is supported by emerging approaches from recent literature that show promise in addressing current limitations.

The remainder of this paper is organized as follows and in Figure 2: Section 2 details our methodology. Section 3 reviews related surveys and positions our work. Section 4 provides a primer on LLMs, including their fundamentals and networking applications. Section 5 explores the integration of LLMs with network technologies like SDN, NFV, IBN, and ZTN. Section 6 presents LLM solutions across four key NO&M domains: Network Design (6.1), Network Automation (6.2), Network Optimization (6.3), and Network Security (6.4). Section 7 discusses challenges and future research directions for LLM-enabled NO&M. Finally, Section 8 concludes with key findings and insights from our survey.

2 Methodology

To ensure a rigorous and transparent survey, we conducted a structured literature review based on a clearly defined methodology consisting of three stages:

(1) **Literature Identification:**
- We searched widely-recognized academic databases including IEEE Xplore, ACM Digital Library, Google Scholar, arXiv.
- Search terms included combinations of relevant keywords: "Large Language Models," "LLMs," "generative AI," "network management," "network design," "network automation," "network security," "network optimization," "GPT," "BERT," "LLaMA," and "computer networking."
- Given the recency of LLM applications in networking (predominantly emerging since 2021), we did not impose a specific start date boundary.
- Our search was limited to English-language publications with an end date of March 2025.
- This initial search yielded 201 potential papers for consideration.

(2) **Literature Screening and Selection:**
- Initial screening was based on title, abstract, and keywords, ensuring relevance to LLM applications in NO&M.
- We established inclusion criteria requiring either substantive conceptual contributions or experimental results/use cases specifically in the NO&M domain.
- Additionally, we performed backward and forward reference searching (citation tracking) on relevant papers to ensure completeness of the literature.
- This screening process narrowed our focus to 67 papers that met our criteria for in-depth analysis. And these papers is just the paper fall in the NO&M, do not include other background, related work, dataset papers.

(3) **Literature Evaluation and Synthesis:**
- Selected papers were categorized based on their primary contribution to NO&M domains: Network Design, Network Automation, Network Optimization, and Network Security.
- This categorization reflects both the core stages in the network management lifecycle and the natural groupings that emerged from the research directions in the collected papers.
- When papers addressed multiple domains, we classified them according to their predominant focus area.
- For each category, papers were analyzed to identify and compare methodologies, challenges addressed, the type of LLM techniques employed (e.g., prompt engineering, fine-tuning), and key findings.
- We evaluated papers based on implementation evidence, experimental validation, and direct relevance to network operations and management.

3 Related Work

The application of LLMs in NO&M has gained increasing attention, with recent studies exploring their impact on various aspects of modern networking. While existing surveys provide valuable insights, their coverage of these topics varies in scope and depth.

Recent works have examined the role of LLMs in shaping the next generation of networking. In telecommunications, studies such as [153] and [14] have explored how these models contribute to network design and automation by enhancing strategic decision-making, optimizing traffic flows, and improving configuration management. Notably, research on ISTN, as discussed in [60], has demonstrated how LLMs facilitate seamless connectivity by optimizing resource allocation across diverse communication infrastructures.

The increasing complexity of modern network environments has also spurred interest in LLMs for network automation. Surveys such as [20] and [131] have highlighted how these models enable real-time analytics, fault detection, and intelligent traffic engineering. As networks continue to evolve toward self-managing architectures, LLMs are being positioned as key enablers of adaptive and autonomous network control mechanisms.

Another area of study involves network optimization, where LLMs have been applied to enhance performance and efficiency in wireless networks. Investigations into AI-driven network intelligence for 6G, such as [23] and [25], have demonstrated that LLMs can support predictive analytics and automated configuration adjustments to optimize resource utilization. Additionally, [68] provides a comprehensive discussion on how AI models, including LLMs, are leveraged for enhancing network optimization and traffic engineering. Moreover, their application in edge computing environments, as seen in [17] and [26], has shown promise in reducing latency and computational overhead, further reinforcing their role in enhancing network efficiency.

Security remains an important concern in NO&M, and LLMs are increasingly being explored for their potential to bolster cybersecurity frameworks. Studies such as [149] has demonstrated how LLMs contribute to intrusion detection, anomaly detection, and threat mitigation strategies. Additionally, systematic reviews of their application in cybersecurity have emphasized their role

Conceptual Map of Survey	
Section 2 Methodology	
Section 3 Related Work	
Section 4 LLM 101	4.1 What Are LLMs
	4.2 LLMs in Networking
	4.3 Workflow of LLMs for Networking
	4.4 LLM Usage in Existing Work
Section 5 LLMs and Network Technologies	5.1 SDN
	5.2 NFV
	5.3 IBN
	5.4 ZTN
	5.5 Comparative Analysis and Future Directions
Section 6 LLM Solutions for NO&M	6.1. Network Design
	6.2. Network Automation
	6.3. Network Optimization
	6.4. Network Security
Section 7 Challenges and Future Directions Of LLM-enabled NO&M	7.1. Challenges and Open Issues
	7.2. Future Perspective and Research Directions
Section 8 Conclusion	

Figure 2: Conceptual map of the survey. The map illustrates the hierarchical structure of the survey, starting from Section 2 (Methodology) through Section 8 (Conclusion).

in automating security protocols and identifying vulnerabilities in complex network infrastructures.

Advancements in mobile edge intelligence have also influenced the deployment and efficiency of LLMs in network management. The survey [107] highlights how edge intelligence frameworks are being leveraged to enhance LLM performance while ensuring low-latency decision-making in distributed networks.

Despite these advancements, gaps remain in existing literature regarding integrating LLMs across all NO&M dimensions. While some surveys provide in-depth discussions on specific topics, few offer a holistic perspective encompassing network design, automation, optimization, and security in a unified framework. Our study aims to bridge this gap by synthesizing recent advancements and presenting a structured analysis of LLM-driven NO&M methodologies. By evaluating the interplay between these four core areas, our work provides a broader and more cohesive understanding of how LLMs can enhance the next generation of intelligent networking systems. A summary of these related works is provided in Table 1.

4 LLM 101

4.1 What Are LLMs?

LLMs have emerged as a transformative technology in the field of AI, demonstrating remarkable capabilities in natural language understanding, generation, and reasoning [22, 114, 141]. These models, trained on vast amounts of textual data, have the potential to revolutionize various domains, including the critical area of networking. The application of LLMs in computer networking holds promise in enhancing network design, automation, security, and others.

LLMs can be categorized into two primary paradigms: open-source and closed-source models, with hybrid approaches emerging as a synthesis of these competing frameworks [88]. Open-source LLMs, such as LLaMA [129], and Mixtral [64], prioritize transparency, collaborative innovation, and architectural modularity by making their model weights and training methodologies publicly accessible. This approach fosters a decentralized research ecosystem that enables external verification, iterative development, and rigorous scrutiny of potential biases [88]. In the networking domain, open-source models offer particular advantages through their parameter-efficient fine-tuning capabilities, which facilitate domain-specific adaptations without prohibitive computational requirements. While historically lagging in performance, recent advancements in optimization techniques such as LoRA and RAG have significantly narrowed the performance gap with proprietary alternatives [88].

Closed-source LLMs, including ChatGPT [103], Gemini [127], and Claude [9], operate within a proprietary framework that restricts external access to underlying architectures and training data [88]. These models leverage substantial computational resources and exclusive datasets, providing a competitive edge in standardized benchmarks while ensuring controlled deployment in high-stakes domains. Proponents argue that this centralized approach mitigates risks associated with adversarial manipulation and enables more robust security protections [88]. However, the opacity of these systems raises concerns regarding accountability, potential algorithmic biases, and limited external scrutiny. Despite their current performance advantages in complex networking tasks, closed-source models face growing pressure to incorporate elements of transparency as regulatory oversight and ethical considerations around AI governance intensify [88].

LLMs are termed 'large' primarily due to the vast number of parameters they contain, often running into hundreds of billions, such as GPT-3, which has 175 billion parameters [108]. These parameters enable LLMs to process and generate language at scale. These models are trained on massive datasets, often containing billions of words or tokens, sourced from diverse domains such as books, websites, and social media [41, 44, 110]. This extensive training allows them to generate content and make predictions by identifying patterns in the data, simulating an understanding of grammar, context, and even subtle nuances like tone and sentiment [16, 151]. The 'model' aspect refers to their ability to generate new content, make predictions, and perform complex tasks based on this pattern recognition.

These are some key concepts of LLMs:

Transformer Architecture: LLMs are mainly built on Transformer architecture, introduced by Vaswani et al. in 2017 [130]. This architecture uses mechanisms like attention to capture long-range dependencies in text, allowing the model to understand relationships between words across entire sentences and paragraphs. Transformers rely on self-attention mechanisms that weigh the importance of different words in a sequence, giving LLMs their remarkable ability to handle context across large spans of text [34, 109].

Pre-training and Fine-tuning: The creation of LLMs generally follows a two-step process: pre-training and fine-tuning. During pre-training, the model learns general language patterns by predicting the next word in a sentence or filling in masked words based on the surrounding context [36, 108]. This phase involves training on vast corpora of data and results in a model that understands general language structure. In the fine-tuning phase, the pre-trained model is adapted to specific tasks or domains by training on smaller, task-specific datasets, which may include labeled data for tasks like translation, summarization, or question-answering [56, 61]. In our survey, most works leverage pre-trained LLMs for their strong generalization capabilities and lower training costs. However, [84] takes a more tailored approach by fine-tuning LLMs specifically for telecommunications. By adapting models like Tinyllama-1.1B [150], Phi-1.5 [75], Gemma-2B, and LLama-3-8B with a domain-specific dataset [85], this work enhances their suitability for telecom-related tasks, addressing industry-specific challenges more effectively.

Tokenization: Tokenization [80] is an essential pre-processing step in LLM training that parses the text into non-decomposing units called tokens. Language is broken down into smaller components, known as "tokens," which can be words, subwords [71], or even characters [117], depending on the model [117]. LLMs operate by predicting and generating sequences of tokens. The choice of tokenization plays a crucial role in how well the model can handle rare or specialized words, making it possible for LLMs to process languages with varying vocabularies and structures [65, 134]. In

Table 1: Survey of LLM applications in networking

Ref.	Year	Focus Area	Network Design	Network Automation	Network Optimization	Network Security	Surveyed Works
[126]	2023	6G Systems	High	High	High	Moderate	152
[14]	2024	Telecom	Moderate	High	Moderate	Low	15
[17]	2024	Edge Computing	Low	Moderate	High	Moderate	45
[23]	2024	6G Wireless	Moderate	Moderate	High	Moderate	303
[153]	2024	Telecom	Moderate	High	High	High	230
[26]	2024	Edge Intelligence	Moderate	High	High	Moderate	15
[60]	2024	Satellite-Aerial-Terrestrial Networks	Moderate	High	High	Moderate	246
[131]	2024	Mobile and Wireless Networking	Moderate	High	Moderate	High	240
[48]	2024	Next-Generation Networking	Moderate	Moderate	High	High	210
[68]	2025	Next-Generation Wireless Networks	Moderate	Moderate	High	High	117
[149]	2025	Cybersecurity	Low	Moderate	Low	High	257
[107]	2025	Mobile Edge Intelligence	Moderate	High	Moderate	Moderate	305
[20]	2025	Network Monitoring and Management	Low	Moderate	Moderate	High	189
[25]	2025	6G	High	High	High	High	15
This Work	**2025**	**LLMs in NO&M**	**High**	**High**	**High**	**High**	**159**

Low: The paper *briefly mentions* or touches on the topic but does not provide in-depth discussion, methods, or case studies. It is not a primary focus.
Moderate: The paper *discusses the topic in some detail*, includes related challenges, some solutions, or comparisons, but it is not the central focus.
High: The paper *directly focuses on this aspect*, provides extensive analysis, methodologies, frameworks, or a dedicated section covering the topic.

networking, tokenization plays a vital role when handling structured data formats like configuration scripts, logs, and command-line instructions. For example, consider this network configuration snippet:

```
interface ethernet0
  ip address 192.168.1.1/24
  no shutdown
```

A simple tokenization would break this down into individual tokens like `interface`, `ethernet0`, `ip`, `address`, `192.168.1.1/24`, `no`, and `shutdown`. This allows LLMs to understand the structure and semantics of the configuration, rather than treating it as unstructured text. Similarly, in log files or command-line output, tokenization helps identify important elements such as timestamps, error codes, hostnames, and other relevant information. This capability enables LLMs to perform tasks like generating network configurations, identifying anomalies in log data, and parsing command-line instructions for automated network management tasks.

Context Window: LLMs operate on a fixed-length context window, meaning they can process a certain number of tokens at a time. While this allows for understanding context within the window, any text exceeding this window length must be truncated or summarized [27]. LLMs have increased context windows, enabling them to manage longer dependencies in text, which is important for tasks like document summarization or coding assistance [22, 54]. For instance, ChatGPT-4 has a context window of 128,000 tokens [104], and Code Llama 70B supports up to 100,000 tokens [113]. In the field of computer networking, the context window is important since network data often includes logs, configurations, and machine-generated text that require larger windows to maintain coherence over long sessions.

Generative Capabilities: One of the powerful features of LLMs is their generative nature. Given a prompt or a starting sentence, they can generate coherent and contextually relevant text. This allows LLMs to perform tasks such as text completion, story generation, or even answering questions in a conversational manner [33, 109]. They are capable of producing human-like responses by leveraging the patterns they have learned during training. In networking, generative capabilities are leveraged to automate tasks like configuration generation, protocol adaptation, and even network troubleshooting, where LLMs suggest possible solutions to network issues based on historical data.

Few-Shot and Zero-Shot Learning: Unlike traditional ML models that require extensive labeled data to perform well on specific tasks, LLMs can excel at few-shot or even zero-shot learning [70, 140]. This means that they can generalize to new tasks with little or no task-specific training [22]. For example, an LLM can perform translation or summarization with just a few examples or even without any prior task-specific training, thanks to the general knowledge it has acquired during pre-training. In the studies we investigated, prompt engineering techniques are extensively used to direct LLMs towards networking tasks, such as generating precise router commands or translating policy intents into network configurations [93]. Studies like the study on Intent-based Networks [42] highlight the effectiveness of carefully crafted prompts in achieving high accuracy with minimal training, which is crucial for dynamically changing networking environments.

Model Scaling and Runtime Efficiency: LLMs demonstrate improved capabilities when trained with increased model size and larger datasets, with models like GPT having tens of billions of parameters [22]. However, this improved capability comes with inference-time challenges. The computational demands of these large models create runtime scalability issues, leading to high infrastructure costs and energy consumption for inference operations

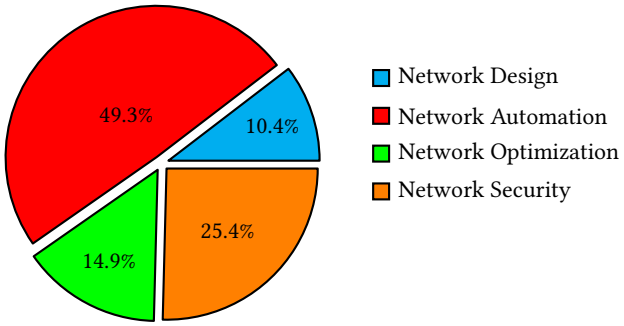

Figure 3: Distribution of existing papers in different networking directions.

[146]. This presents a particular challenge for networking applications, where real-time performance and low latency are essential requirements. Effective networking solutions must balance the capabilities of these powerful models with practical deployment constraints, potentially through techniques like model compression, distillation, or specialized hardware to achieve acceptable inference efficiency on existing infrastructure [79].

Why Are LLMs Important in Networking? LLMs represent a shift in how machines interact with both structured and unstructured data [7]. Their ability to generate human-like text and analyze machine-generated data has transformed industries such as healthcare [30], legal analysis, content creation, and customer support. In the realm of computer networking, LLMs can automate and enhance tasks traditionally requiring human expertise, such as protocol configuration, dynamic resource allocation, and anomaly detection. Their adaptability and capability to handle diverse data sources make them promising tools with potential for managing modern, complex network infrastructures.

4.2 LLMs in Networking

LLMs are being applied in various aspects of networking, offering solutions to complex tasks traditionally requiring human expertise. The primary areas where LLMs are making an impact include **Network Design**, **Network Automation**, **Network Optimization**, and **Network Security**. In the section 6, we will present the existing works in these areas.

Figure 3 shows the distribution of existing papers in the areas of Network Design, Network Automation, Network Optimization, and Network Security, based on the 67 papers we analyzed. As illustrated, Network Automation accounts for 49.3% of the contributions, followed by Network Security at 25.4%. This demonstrates the main focus on using LLMs to automate and streamline operational tasks while maintaining secure and optimized network environments.

4.3 Workflow of LLMs for Networking

The workflow of LLMs in networking involves multiple stages, each designed to harness the capabilities of LLMs for automating and optimizing network management and operation tasks. This process integrates various techniques and models, ensuring that LLMs perform efficiently across a range of tasks such as configuration management, anomaly detection, and network optimization. Below, we present a structured workflow that draws from recent studies and is based on the workflow proposed by Liu et al. [79] to capture how LLMs are applied in networking environments. This workflow includes seven key stages: task definition, data representation, prompt engineering, model training, tool integration, validation, and deployment as shown in Figure 4. While these stages are interconnected, they are not fully coupled, allowing for flexibility and adaptability in the implementation process.

- **Task Definition**: This stage involves specifying the network management task that LLMs will perform. Tasks could range from automated configuration and protocol synthesis to incident detection and performance optimization. Each task must be clearly defined to guide the LLM toward generating actionable outputs. For instance, studies like [96] emphasize the importance of defining tasks such as troubleshooting network issues and automating administrative tasks clearly before feeding the models.
- **Data Representation**: Data representation is important to ensure that the LLM can process and understand network-specific data, which may include network logs, routing tables, traffic data, and configurations. The preprocessing step typically involves converting this data into formats that LLMs can process effectively. Tokenization plays a important role here, as LLMs break down input data into smaller units such as tokens, words, or phrases [80, 106]. Studies like PROSPER [121] and NetConfEval [132] demonstrate how LLMs can handle diverse network data formats, making this preprocessing step essential for success.
- **Prompt Engineering**: Prompt engineering is the process of crafting prompts that guide the LLM in generating relevant and accurate outputs [115, 144]. This is a crucial step for LLMs to perform well in network tasks like configuration generation and policy translation. Studies such as [38] show the importance of designing prompts that lead the model to generate desired results. Techniques like few-shot learning [139], chain-of-thought prompting [142] and contextualizing inputs are applied to ensure the LLM produces accurate, reliable responses to network tasks [22, 70, 142].
- **Model Training**: While many existing systems rely on pre-trained models like GPT-4 or LLaMA, fine-tuning these models with network-specific datasets can enhance their performance. Studies such as PreConfig [73] illustrate the value of fine-tuning LLMs to specific network environments, thereby improving their accuracy in handling configurations or detecting network anomalies. For instance, PreConfig achieved impressive results across several key tasks after fine-tuning, including a BLEU score of 82.25% in configuration generation, 95.76% in configuration analysis, and 89.2% in configuration translation, showcasing its enhanced ability to handle network-specific tasks with high precision. Fine-tuning allows the model to adapt to the unique data and tasks present in networking, which enhances its ability to generate precise outputs.

- **Tool Integration**: LLMs do not operate in isolation; they are often integrated with external tools like network simulators, verification systems, and monitoring platforms. For example, in AppleSeed [77], LLMs are integrated with multi-domain infrastructure management tools to translate configuration tasks into action across different vendors. Similarly, external tools like Batfish [15] are commonly integrated for tasks such as configuration verification or network performance simulation, ensuring the LLM's recommendations are accurate before being applied.
- **Validation**: Validation is crucial in ensuring that the LLM's outputs are accurate and reliable. In NetLM [147] and [21], validation mechanisms involve continuous testing in sandbox environments or real-time monitoring to ensure the model's suggestions are correct before deployment.
- **Deployment**: Once the LLM is validated, the model is deployed into the live network environment. This deployment could involve real-time automation tasks such as dynamic traffic routing, configuration changes, or incident management. Studies such as [92] emphasize that real-time execution enables adaptive network behaviors, where the LLM continually adjusts configurations based on network changes and user intent. Furthermore, incorporating a feedback loop allows for continuous learning, where the model is retrained or adjusted based on the performance of its recommendations.

By following this workflow, LLMs can be effectively integrated into a wide range of networking tasks, from automating network configurations and optimizing performance to improving security and incident detection. The adaptability of LLMs in each stage of the workflow ensures that they can operate efficiently in dynamic and complex network environments, providing improvements in both automation and optimization.

4.4 LLM Usage in Existing Work

Various LLMs are currently used across different networking tasks, with **GPT-based models** dominating the field. This analysis is based on the papers we reviewed that explicitly implemented LLMs in their work, excluding those that only discuss LLMs theoretically. From this review, GPT-based models from OpenAI constitute the majority of usage, with **GPT-3.5** being the most widely adopted model, accounting for approximately 22.73%, followed closely by **GPT-4** at 20.91%. Additionally, the OpenAI models collectively represent the largest share, with a **52.73%** provider usage across the reviewed papers, highlighting their extensive adoption for handling complex networking scenarios, such as traffic analysis and configuration generation.

Apart from OpenAI's models, **Meta's LLaMA family** also demonstrates big usage, with a combined provider usage of 16.36%, primarily through models like **LLaMA-2** and **LLaMA-3**. Other models, such as **CodeLLaMA**, **Gemini**, and **Mistral**, are less prevalent in comparison.

Our decision to group these LLMs into families, rather than listing each variant separately, is to provide clearer insights into model preferences without unnecessary clutter. Grouping by family (e.g., GPT, LLaMA, Gemini) allows us to showcase the dominance of certain LLMs across networking applications without overwhelming the analysis with individual model variations.

For a detailed breakdown of models included in each LLM family, see Table 2.

5 LLMs and Network Technologies

The evolution of NO&M has progressed from manual processes to intelligent, autonomous systems driven by increasing network complexity and demands for efficiency. The integration of LLMs with network paradigms like SDN [57], NFV [102], IBN [29], and ZTN [43] is reshaping modern NO&M by enhancing automation, decision-making, and real-time adaptation capabilities. In this section, we specifically focus on papers that explicitly demonstrate LLM integration with these network technologies, rather than those that discuss theoretical applications or general network operations. This approach ensures our analysis accurately represents the current state of practical integration. This section explores how LLMs complement these technologies, examining current contributions.

5.1 SDN

SDN centralizes network control by decoupling the control and data planes, allowing for dynamic and programmable network management. LLMs can enhance SDN through policy automation and dynamic control, where they assist in generating and enforcing network rules.

Research by Li et al. [76] demonstrates how LLMs can enhance SFC deployment and optimization within SDN-enabled environments. Their NSGA2-based multi-objective optimization algorithm leverages LLMs to provide heuristic functions for SFC deployment, outperforming both model-based and greedy algorithms in QoS metrics while minimizing payment costs. This approach shows how LLMs can transform SDN's programmable nature by automatically generating network configurations based on high-level requirements.

5.2 NFV

NFV enables the deployment of network services as virtualized functions, reducing dependence on specialized hardware. LLMs have demonstrated potential to enhance NFV environments in several key areas.

Mekrache et al. [92] presented an LLM-based Intent translation system that converts natural language expressions into NSDs. Their approach employs few-shot learning on an open-source LLM with a Human Feedback loop, achieving high accuracy in generating NSDs ready for deployment on edge computing clusters. This implementation demonstrates how LLMs can automate service deployment and orchestration by translating operator intents into NFV SFCs.

Network slicing, a application of NFV in 5G/6G networks, has also benefited from LLM integration. Bandara et al. [12] developed SliceGPT, which employs a fine-tuned GPT-3.5-turbo model with blockchain and smart contracts to optimize network slice orchestration. Their system supported over 500 transactions per second for network slice trading, demonstrating practical scalability for

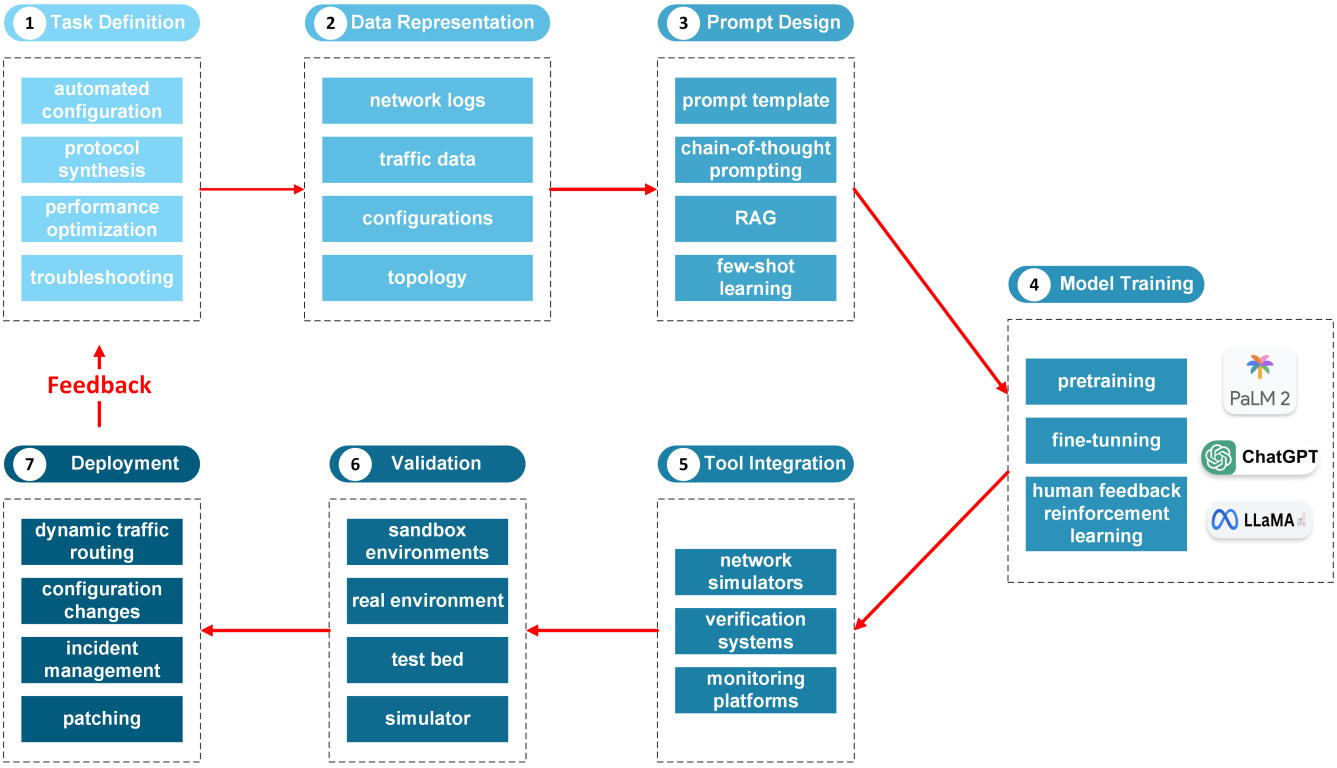

Figure 4: A workflow of LLMs for networking. The workflow illustrates a seven-stage process: Task Definition, Data Representation, Prompt Design, Model Training (utilizing PaLM 2, ChatGPT, and LLaMA), Tool Integration, Validation, and Deployment. Each stage contains specific components for network operations, forming a continuous improvement cycle with feedback mechanisms.

real-world deployment. This integration enables transparent, efficient resource sharing while optimizing slice configurations based on specific application requirements.

Ghasemirahni et al. [45] introduced FlowMage, a system that leverages LLMs to perform code analysis and extract essential information from stateful network functions prior to deployment. By analyzing NF code with GPT-4, FlowMage finds optimized configurations when deploying stateful NF chains, resulting in significant performance improvements (up to 11×) compared to default configurations. This demonstrates how LLMs can enhance the resource efficiency of virtualized network functions.

5.3 IBN

IBN shifts network management toward intent-driven operations, where users specify high-level objectives instead of low-level configurations. LLMs are particularly well-suited for IBN due to their natural language processing capabilities, and our analysis shows IBN has the highest number of LLM integration papers among network technologies.

LLMs can enhance IBN in two areas: intent fulfillment and intent assurance. For intent fulfillment, Mekrache et al. [93] introduced an LLM-centric Intent Life-Cycle management architecture that leverages CodeLlama to transform natural language network requests into actionable infrastructure-level configurations. Their comprehensive framework addresses the complete intent lifecycle—from decomposing multi-domain intents and translating them into domain-specific configurations to handling negotiation, activation, and assurance procedures. In user evaluations conducted with 10 volunteers at EURECOM, their system achieved an impressive initial rating score of approximately 4.5 out of 5 for intent decomposition capability, demonstrating the effectiveness of LLMs in understanding complex networking requirements expressed in natural language.

Similarly, Guo et al. [46] proposed LLM-empowered Intent Translation, which combines GPT-4 with RAG and MoE to generate high-quality network policies. This approach demonstrated a 56.7% improvement in F1 score compared to baseline models, with precision, recall, and F1 score improvements of 35.54%, 37.61%, and 36.55% respectively after implementing RAG.

For intent assurance, LLMs contribute by continuously monitoring intent execution and making real-time adjustments to maintain policy compliance. [11] combined advanced LLMs with KNN classification for intent translation and contradiction detection, achieving up to 5% higher accuracy (88%) and improved F1 scores compared to existing methods.

Table 2: Distribution of LLM Usage in Reviewed Papers

Provider	LLM Family	Model Family	Specific Models	Source Type	Model Family Usage (%)	Provider Usage (%)
OpenAI						52.73%
	GPT [22]	GPT-2	GPT-2	Closed	2.73%	
		GPT-3	GPT-3	Closed	0.91%	
		GPT-3.5	GPT-3.5, text-davinci-003, GPT-3.5-turbo, ChatGPT (GPT-3.5 based)	Closed	22.73%	
		GPT-4	GPT-4, GPT-4-turbo, GPT-4o	Closed	20.91%	
		ChatGPT	ChatGPT (No specific version)	Closed	5.45%	
Meta						16.36%
	LLaMA [129]	LLaMA	No specific version	Open	0.91%	
		LLaMA-2	LLaMA-2: 7B, 13B, chat	Open	6.36%	
		LLaMA-3	LLaMA-3/3.1: 7B, 8B, 70B, 405B	Open	5.45%	
		CodeLLaMA	CodeLLaMA: 7B, 13B, 34B	Open	3.64%	
Google						14.55%
	Bard [34]	Bard	Bard	Closed	1.82%	
	BERT [34]	BERT	BERT	Open	7.27%	
	Gemini [127]	Gemini	Gemini-1.0-Pro, Bard	Closed	1.82%	
		Gemma	Gemma: 2B, 7B	Open	1.82%	
	T5 [110]	T5	Multimodal-CoT-T5: 223M, 738M	Open	1.82%	
Anthropic						1.82%
	Claude [8]	Claude	Claude-2, Claude-3	Closed	1.82%	
Mistral AI						2.73%
	Mistral [63]	Mistral	Mistral-7B	Open	2.73%	
TII						0.91%
	Falcon [6]	Falcon	Falcon	Open	0.91%	
Alibaba						1.82%
	Qwen [2]	Qwen	Qwen-4B	Open	1.82%	
Various						9.09%
	Others	Others	BASH-Coder-Mistral-7B, Zephyr-7B-β, TeleRoBERTa, Bing Copilot, Github Copilot, Baichuan, ChatGLM, ALBERT, Moonshot, Vicuna, Phi2	Mixed	9.09%	

Note: TII stands for Technology Innovation Institute

The IBN domain shows notable differences in performance between open-source and closed-source LLMs. Fuad et al. [42] highlighted that while GPT models can generate correct router BGP configurations, open-source models like Llama2 and Mistral showed inconsistencies requiring additional tuning. Complex configurations such as firewalls posed challenges for all tested models, necessitating few-shot learning and prompt engineering to generate correct results.

5.4 ZTN

ZTN represents the next step toward full automation in network management by eliminating manual interventions. LLMs play a crucial role in enabling zero-touch networking through autonomous configuration and self-healing capabilities.

Lira et al. [78] proposed LLM-NetCFG, which employs Zephyr-7B-β to automatically generate, verify, and deploy network configurations based on natural language intents with minimal human intervention. This approach demonstrates how LLMs can enable zero-touch provisioning by generating, validating, and applying network settings autonomously.

Ali et al. [4] explored the integration of LLMs into ZTNs for DRL-based anti-jamming strategies, highlighting how LLMs can distill intricate network operations into intuitive, human-readable reports while enhancing automated decision-making. Their approach bridges the gap between automated processes and human-centric interfaces, improving transparency in autonomous networks.

Mekrache et al. [94] introduced a novel pipeline for ensuring trustworthy ZSM in 6G networks by combining AI for detecting anomalies, XAI to identify root causes using feature importance analysis, and LLMs to generate user-friendly explanations and suggest corrective actions. Their framework demonstrated efficiency in scaling cloud resources to prevent SLA violations while providing understandable explanations, enhancing trust in autonomous systems.

5.5 Comparative Analysis

Our analysis reveals varying levels of integration maturity across different network technologies. IBN shows the highest level of LLM integration, likely due to the natural fit between LLMs' language capabilities and intent translation. NFV and ZTN demonstrate moderate integration, while SDN applications remain somewhat limited despite the technology's programmable nature.

Table 3 provides a comparative summary of how different LLM capabilities align with each network technology:

As networking technologies continue to evolve, LLMs offer promising enhancements by improving automation, security, and decision-making across SDN, NFV, IBN, and ZTN. While some of these integrations are well-documented in existing research, others remain underexplored, presenting opportunities for future studies.

6 LLM Solutions for NO&M

6.1 Network Design

LLMs are transforming network design by enhancing configuration automation, protocol understanding, and intelligent optimization.

Table 3: Comparative Analysis of LLM Integration with Network Technologies

Technology	Primary LLM Capabilities Leveraged	Current Integration Maturity
SDN	Flow rule optimization, network orchestration automation	Low
NFV	Service orchestration, Resource allocation	Medium
IBN	Intent translation, Validation	High
ZTN	Autonomous configuration, Self-healing	Medium

The integration of verification mechanisms, multimodal capabilities, and domain-specific adaptation enables LLMs to address key challenges in modern network environments.

Recent research by Mondal et al. [98] shows promising results in using LLMs to generate router configurations. Their study found that while GPT-4 alone produces configurations with errors, combining it with automated verification tools dramatically improves results. Their approach, called 'Verified Prompt Programming,' uses specialized verifiers to detect errors and automatically provide targeted feedback to GPT-4, creating a correction loop. This method reduced human intervention by up to 90% compared to manually correcting LLM outputs, with their system handling 10 automated corrections for every human prompt when translating between router configuration languages. These results demonstrate that while LLMs have potential for network automation, they require integrated verification mechanisms to produce reliable configurations. NETBUDDY [133] further enhances configuration efficiency by breaking complex tasks into manageable components with self-healing code generation, achieving 100% translation accuracy for moderate-complexity configurations and demonstrating that batching 10 requirements is 6× more cost-efficient than individual processing.

Protocol understanding also benefits from LLM capabilities. PROSPER [121] employs GPT-3.5-turbo to extract protocol specifications from RFC documents by combining diagram artifacts with textual descriptions. This technique achieves 1.3× more true positives and 6.5× fewer false positives than traditional FSM extraction approaches, improving protocol extraction accuracy. Meanwhile, GeNet [59] leverages GPT-4's multimodal capabilities to process network diagrams alongside textual requirements, enabling comprehensive topology and configuration management from visual inputs.

Domain-specific adaptation enhances LLMs' network intelligence capabilities. WiLL [13] implements RAG with specialized chunking strategies for FCC compliance verification, achieving 78.57% accuracy compared to off-the-shelf models (GPT-4: 73.20%, ChatGPT: 46.42%, LLaMA: 35.71%). This 51.8% improvement over baseline LLM performance demonstrates the importance of domain knowledge integration for regulatory compliance tasks.

Network algorithm optimization represents another advancement. He et al. [51, 52] developed complementary frameworks, Nada and LLM-ABR, that employ Chain-of-Thought prompting to autonomously design network algorithms. These frameworks achieve compilation success rates of 68.6% (GPT-4) versus 41.2% (GPT-3.5) and consistently outperform traditional algorithms across diverse network environments including broadband, satellite, 4G, and 5G.

Table 4 presents quantitative comparisons that reveal the advantages of LLM-based approaches over baseline methods. As shown in the table, LLM solutions provide improvements across multiple dimensions of network design, from protocol extraction and router configuration to regulatory compliance and algorithm optimization.

However, important technical limitations persist, including model performance variability, complexity constraints in configuration tasks, verification requirements for raw outputs, and limitations in current multimodal capabilities.

These advancements demonstrate LLMs' potential to transform network design by automating complex configuration tasks, enhancing protocol understanding, and enabling more intelligent optimization—ultimately making network design more efficient, accurate, and accessible. Table 5 summarizes key research papers focused on LLM applications in network design, highlighting their publication year and primary LLM models used.

6.2 Network Automation

LLMs are transforming network automation by enhancing intent translation, streamlining configuration management, and enabling zero-touch operations. Recent research demonstrates advancements in applying LLMs to automate complex networking tasks, addressing key challenges such as configuration complexity, cross-vendor compatibility, and dynamic network environments.

Intent-based management represents a contribution of LLMs to network automation. Mekrache et al.[93] introduced an LLM-centric Intent Life-Cycle management architecture using CodeLlama to translate natural language requests into actionable network configurations. Their framework encompasses the complete intent lifecycle from decomposition and translation to negotiation, activation, and assurance, achieving an initial rating score of approximately 4.5 for intent decomposition. Similarly, Guo et al.[46] proposed LLM-empowered Intent Translation, which combines GPT-4 with RAG and MoE to generate high-quality network policies. This approach demonstrated a 56.7% improvement in F1 score compared to baseline models, with precision, recall, and F1 score improvements of 35.54%, 37.61%, and 36.55% respectively after implementing RAG.

Cross-vendor configuration translation benefits from LLM capabilities. Wei et al.[143] developed an IRAG framework that systematically splits configuration files into fragments, extracts intents, and generates accurate translations between vendor-specific device systems. Their approach achieved 97.74% syntax correctness, outperforming the GPT-4o baseline by 19.07%. Similarly, Lin et al.[77] introduced AppleSeed, which leverages few-shot learning with ChatGPT to translate intents into intermediate programs for multi-domain infrastructure management. This system achieved

Table 4: Quantitative comparison of LLM-based approaches methods in Network Design

Approach	Baseline Method Performance	LLM Method Performance	Improvement
Protocol Extraction [121]	Baseline false positive rate	6.5× reduction in false positives, 1.3× increase in true positives	Significant reduction in extraction errors
Router Configuration [98]	Manual configuration required	10× leverage for Juniper translation	90% reduction in manual effort
Regulatory Compliance [13]	Human experts required	78.57% accuracy vs. 51.77% for off-the-shelf LLMs	51.8% improvement over baseline LLMs
Configuration Translation [133]	Complex manual translation	100% accuracy for ≤40 requirements; 6× cost efficiency with batching	Perfect accuracy for moderate complexity tasks
ABR Algorithm Design [52]	Default ABR algorithms	Consistent outperformance across diverse network conditions	Performance gains across all network types

Table 5: Papers on LLM Solutions for Network Design

Title	Ref	Year	Primary LLM
Can we make FCC Experts out of LLMs?	[13]	2025	ChatGPT, GPT-4, LLaMa
Designing Network Algorithms via Large Language Models	[51]	2024	GPT-3.5, GPT-4
What do LLMs need to Synthesize Correct Router Configurations?	[98]	2023	GPT-4
PROSPER: Extracting Protocol Specifications Using Large Language Models	[121]	2023	GPT-3.5-turbo
LLM-ABR: Designing Adaptive Bitrate Algorithms via Large Language Models	[52]	2024	GPT-3.5, GPT-4
GeNet: A Multimodal LLM-Based Co-Pilot for Network Topology and Configuration	[59]	2024	GPT-4
Making Network Configuration Human Friendly	[133]	2023	GPT-4

efficient intent translation with an average 22.3x lines of code to intent word ratio, while also speeding up management plan execution by 1.7-2.6 times through just-in-time compilation.

Zero-touch network management advances through LLM integration. Lira et al.[78] proposed LLM-NetCFG, which employs Zephyr-7B-β to automatically generate, verify, and deploy network configurations based on natural language intents with minimal human intervention. Jeong et al.[62] developed S-Witch, a system using GPT-3.5-turbo and NDT technology to convert user requirements expressed in natural language into CLI commands for traditional network switches, making automation accessible even in environments with legacy equipment.

Domain-specific adaption enhances LLM capabilities for specialized networking tasks. Kan et al. [67] developed Mobile-LLaMA, a specialized version of LLaMA 2 13B trained specifically for 5G network analysis tasks. Unlike general-purpose LLMs, their model was fine-tuned with network-specific instructions and demonstrated performance improvements in analyzing network data. When evaluated against a custom benchmark covering IP routing, packet analysis, and performance evaluation tasks, Mobile-LLaMA successfully completed 82% of the test cases (scoring 247 points on a 300-point scale), substantially outperforming both standard LLaMA models and GPT-3.5 in IP routing analysis tasks. This demonstrates how domain-specific training can dramatically enhance LLM capabilities for specialized networking operations. Fang et al.[39] developed LLMNDC, a fine-tuned LLM approach for network device configuration using Baichuan2-13B, GLM4-9B, and other models. By combining fine-tuning with local knowledge bases, they achieved a 25% reduction in configuration time and a 30% decrease in error rates compared to traditional approaches.

Semantic routing and classification enhance intent management in complex networks. Manias et al.[90] implemented semantic routing to improve LLM performance in 5G core networks, demonstrating that routing architectures enhance accuracy and efficiency compared to standalone LLM prompting methods. The routing approach also addressed LLM hallucination issues, providing more deterministic and reliable performance over extended use. Asif et al.[11] combined advanced LLMs with KNN classification for intent translation and contradiction detection, achieving up to 5% higher accuracy (88%) and improved F1 scores compared to existing methods.

NDTs provide valuable verification environments for LLM-generated configurations. Zhou et al.[152] introduced Hermes, a framework that uses LLM agents (GPT-4o, LLaMA-3.1) to construct NDT instances through structured blueprints. This approach achieved up to 80% accuracy in network modeling tasks while providing explainable, logical steps for verification. This integration of LLMs with NDTs represents an advancement toward fully autonomous network operations by ensuring generated configurations can be validated before deployment.

Multi-agent collaboration enhances decision-making in complex network environments. Shokrnezhad et al.[123] proposed the Autonomous Reinforcement Coordination framework, which combines LLM-based RAG with Hierarchical Action Planning. Their two-tier approach uses LLaMA 3.1-8B for high-level planning and Reinforcement Learning agents for low-level decision-making, demonstrating effective resource orchestration in SAGINs. Similarly, Brodimas et al.[21] leveraged multimodal generative AI to establish a translation pipeline for network configuration, effectively bridging the gap between user-friendly intent expression and underlying traditional network domains.

Table 6: Quantitative comparison of LLM-based approaches in network automation

Automation Approach	Baseline Method Performance	LLM Method Performance	Improvement
Intent Translation [46]	Baseline ChatGLM-6B model	LIT framework with GPT-4 and RAG	56.7% improvement in F1 score; 35.54% in precision
Cross-Vendor Configuration [143]	Standard GPT-4o	IRAG framework with fine-tuned models	19.07% higher syntax correctness (97.74% total)
Network Device Configuration [39]	Manual configuration	Fine-tuned LLM + knowledge base	25% reduction in configuration time; 30% decrease in error rate
Network Analysis [67]	GPT-3.5 (score: 209/300)	Domain-specific Mobile-LLaMA	Higher score (247/300); superior performance in IP routing analysis
Intent Management [92]	Traditional configuration interfaces	Natural language intent expression	reduced configuration complexity with rating scores 4.5/5
Intent Validation [11]	Existing intent translation methods	LLM + KNN classifier approach	5% higher accuracy (88%) and improved F1 scores

Despite these advances, several challenges remain in LLM-based network automation. Fuad et al.[42] highlighted that while GPT models can generate correct router BGP configurations, open-source models like Llama2 and Mistral showed inconsistencies requiring additional tuning. Complex configurations such as firewalls posed challenges for all tested models, necessitating few-shot learning and prompt engineering to generate correct results. Additionally, Mekrache et al.[93] observed that end-to-end processing time exceeded 2 minutes when translating requests containing more than 3 Cloud/Edge applications, indicating potential scalability concerns for complex deployments.

As shown in Table 6, LLM-based approaches consistently outperform baseline methods across various domains. These advancements demonstrate LLMs' potential to transform network automation by enhancing configuration accuracy, reducing manual effort, and enabling more intuitive intent-based management—ultimately making network operations more efficient, reliable, and accessible to a broader range of users without specialized networking expertise. Table 7 presents research papers exploring LLM solutions for network automation, which represents the largest portion (49.3

6.3 Network Optimization

LLMs offer solutions for network optimization by enhancing resource management, enabling intelligent decision-making, and supporting dynamic adaptability in complex networking environments. Recent research demonstrates advances across various optimization domains including wireless networks, cloud environments, and next-generation mobile systems.

Intelligent resource allocation represents a contribution of LLMs to network optimization. Lee et al.[72] introduced a novel knowledge-free network management paradigm using GPT-3.5-Turbo, which eliminates dependency on scenario-specific information. This approach achieved near-optimal energy efficiency after just 100 iterations through a reward-based self-improvement mechanism and multi-LLM collaboration. Similarly, Du et al.[37] developed an LLM-enabled MoE framework that dynamically selects specialized DRL models based on user requirements. Their approach maintained task completion rates above 85%, demonstrating superior decision-making compared to conventional gate networks without LLM integration. This reduces the need to train new models for each optimization problem, thereby decreasing computational costs while improving performance.

Game theory integration with LLMs provides another powerful optimization approach. He et al.[50] proposed a RAG framework that combines LLMs with game theory for mobile networking optimization. Their system matched expert-designed strategies in UAV secure communication scenarios while outperforming benchmark algorithms. This integration enables automated Nash equilibrium calculations and strategy formulation without requiring extensive domain expertise, making sophisticated game-theoretical approaches more accessible for practical network optimization.

Graph-based optimization benefits from LLM integration. Sun et al.[124] demonstrated that combining LLMs with GNNs creates a more efficient and stable model for UAV trajectory and communication resource allocation compared to alternative approaches like LLM+GAT and LLM+Node2Vec. This hybrid approach leverages the LLM's reasoning capabilities alongside the GNN's structural understanding to optimize dynamic networking parameters more effectively than either technology alone.

Multi-objective optimization represents another advancement. Li et al.[76] introduced an NSGA2-based multi-objective optimization algorithm leveraging LLMs to provide heuristic functions for SFC deployment. Their approach outperformed both model-based and greedy algorithms in QoS metrics while minimizing payment costs. This demonstrates LLMs' ability to balance competing objectives in complex network optimization scenarios, providing more nuanced and efficient solutions than traditional approaches.

Network slicing optimization benefits from LLM capabilities in 5G/6G environments. Bandara et al. .[12] created SliceGPT, an innovative marketplace that combines GPT-3.5-turbo with distributed ledger technology to forecast network demands and automate resource allocation. Their decentralized platform processes more than 500 transactions per second in network slice trading scenarios, proving its viability for production environments. By encoding network slices as non-fungible tokens and leveraging GPT-3.5's analytical capabilities, the system optimizes slice configurations dynamically based on application-specific requirements while ensuring fair distribution of resources among service providers.

Table 7: Papers on LLM Solutions for Network Automation

Title	Ref	Year	Primary LLM
Adaptive Resource Allocation Optimization Using Large Language Models in Dynamic Wireless Environments	[101]	2025	GPT-3.5 Turbo
Deploying Stateful Network Functions Efficiently using Large Language Models	[45]	2024	GPT-4 Turbo, GPT-3.5 Turbo, CodeLlama-34B-Instruct, Gemini-1.0-Pro
Evaluating Large Language Models for Optimized Intent Translation and Contradiction Detection Using KNN in IBN	[11]	2025	BERT-bu, GPT2, LLaMA3, Claude2
Following the Compass: LLM-Empowered Intent Translation with Manual Guidance	[46]	2024	GPT-4, ChatGLM-6B, Baichuan-13B-LoRA
Generative AI-in-the-loop: Integrating LLMs and GPTs into the Next Generation Networks	[148]	2024	GPT-3.5
Leveraging Large Language Models for DRL-Based Anti-Jamming Strategies in Zero Touch Networks	[4]	2023	Falcon 7B
Leveraging LLM Agents for Translating Network Configurations	[143]	2025	GPT-4o, Qwen-Max
LLM-Based Intent Processing and Network Optimization Using Attention-Based Hierarchical Reinforcement Learning	[47]	2024	BERT, ALBERT
LLMNDC: A Novel Approach for Network Device Configuration based on Fine-tuned Large Language Models	[39]	2024	Baichuan2-13B, GLM4-9B, LLaMA2, Qwen2
NetOrchLLM: Mastering Wireless Network Orchestration with Large Language Models	[1]	2024	Mistral, ChatGPT 4
On Combining XAI and LLMs for Trustworthy Zero-Touch Network and Service Management in 6G	[94]	2024	Llama2, CodeLlama, Vicuna, Phi2, Gemma
RAG-inspired Intent-Based Solution for Intelligent Autonomous Networks	[100]	2025	GPT-4, BERT
LLM-enabled Intent-driven Service Configuration for Next Generation Networks	[92]	2024	Code Llama
Mobile-LLaMA: Instruction Fine-Tuning Open-Source LLM for Network Analysis in 5G Networks	[67]	2024	GPT-3.5, LLaMA
Intent-Based Management of Next-Generation Networks: an LLM-centric Approach	[93]	2024	CodeLlama
Enhancing Network Management Using Code Generated by Large Language Models	[89]	2023	GPT-4, GPT-3, text-davinci-003, Google Bard
WirelessLLM: Empowering Large Language Models Towards Wireless Intelligence	[120]	2024	GPT-3.5, GPT-4, Claude-3 Opus
Network Meets ChatGPT: Intent Autonomous Management, Control and Operation	[135]	2023	ChatGPT
Towards Intent-based Network Management for the 6G System adopting Multimodal Generative AI	[21]	2024	GPT-3.5-turbo
An Intent-based Networks Framework based on Large Language Models	[42]	2024	GPT-4, GPT-3.5, Llama-2-7B, Mistral-7B
Large Language Models meet Network Slicing Management and Orchestration	[31]	2024	–
LLM-based policy generation for intent-based management of applications	[38]	2023	GPT-3.5, GPT-4
S-Witch: Switch Configuration Assistant with LLM and Prompt Engineering	[62]	2024	GPT-3.5-turbo
Semantic Routing for Enhanced Performance of LLM-Assisted Intent-Based 5G Core Network Management and Orchestration	[90]	2024	Mistral-7B-Instruct-v0.2
AppleSeed: Intent-Based Multi-Domain Infrastructure Management via Few-Shot Learning	[77]	2023	ChatGPT
Telco-RAG: Navigating the Challenges of Retrieval Augmented Language Models for Telecommunications	[19]	2024	GPT-3.5
Towards Intent-Based Network Management: Large Language Models for Intent Extraction in 5G Core Networks	[91]	2024	GPT-3.5
Large Language Models for Zero Touch Network Configuration Management	[78]	2024	Zephyr-7B-β
Hermes: A Large Language Model Framework on the Journey to Autonomous Networks	[152]	2024	GPT-4o, LLaMA-3.1-70B, LLaMA-3.1-405B
Automation of Network Configuration Generation using Large Language Models	[24]	2024	GPT-3.5-Turbo, Llama-2-7B
An Autonomous Network Orchestration Framework Integrating Large Language Models with Continual Reinforcement Learning	[123]	2025	LLaMA 3.1–8B
Adapting Network Information to Semantics for Generalizable and Plug-and-Play Multi-Scenario Network Diagnosis	[125]	2025	GPT-4o, GPT-3.5

Adaptability to diverse network conditions represents a key advantage of LLM-based optimization. Wu et al.[147] proposed NetLLM, which adapts Llama2-7B through LoRA-based fine-tuning for various networking tasks. Their approach outperformed state-of-the-art algorithms, reducing mean absolute error by 10.1-36.6% for viewport prediction, improving Quality of Experience by 14.5-36.6% for adaptive bitrate streaming, and reducing job completion time by 6.8-41.3% for cluster job scheduling. This demonstrates LLMs' ability to serve as foundation models across diverse optimization tasks without requiring task-specific model design.

Despite these advances, technical limitations remain. Performance can vary based on the complexity of optimization problems, and some LLM-based approaches require significant computational resources during training. Additionally, real-time optimization in highly dynamic networks may present latency challenges that future research must address. As shown in Table 8, LLM integration into network optimization shows promising results compared to traditional approaches in specific scenarios. The data indicates performance improvements in areas such as knowledge-free resource management, multi-expert optimization, and task-adaptive networking, though performance varies across different optimization tasks and environments. These advancements illustrate LLMs' potential to transform network optimization by reducing the need for domain-specific knowledge and custom model development while improving performance across diverse networking environments. As LLM capabilities continue to evolve, their integration into network optimization frameworks promises increasingly intelligent, efficient, and adaptive networking solutions. Table 9 provides an overview of papers applying LLMs to network optimization challenges across resource management, multi-objective optimization, and adaptive network systems.

6.4 Network Security

LLMs offer innovative solutions for network security challenges by enhancing threat detection capabilities, improving explainability of security decisions, enabling privacy-preserving operations, and facilitating automated vulnerability testing. Recent research demonstrates advancements in applying LLMs across various security domains, with particular emphasis on their ability to provide more accurate, explainable, and privacy-conscious security solutions.

Anomaly and intrusion detection represents a core strength of LLM-based security approaches. Maasaoui et al.[83] employed BERT with a specialized Byte-level Byte-pair Encoding tokenizer to identify network-based attacks in IoT environments, outperforming traditional classification methods across multiple datasets. Similarly, Ferrag et al.[40] introduced SecurityBERT, which combines BERT with a novel Privacy-Preserving Fixed-Length Encoding technique, achieving an impressive 98.2% accuracy in classifying fourteen distinct attack types while maintaining a compact 16.7MB model size suitable for resource-constrained IoT devices. These advancements demonstrate LLMs' ability to detect sophisticated attacks with greater accuracy than conventional methods while addressing the privacy and resource constraints inherent in security applications.

Network traffic classification benefits from LLM capabilities. Zhou et al.[154] utilized fine-tuned GPT models for classifying network traffic, achieving a mean accuracy of 94.84% across diverse datasets, outperforming traditional ML methods like Naïve Bayes, SVM, and MLP. Their approach demonstrates that proper fine-tuning of LLMs can improve classification accuracy from 43.30% (base GPT-3.5-Turbo) to 94.84% (fine-tuned model), highlighting the importance of domain-specific adaptation. This performance improvement enables more accurate identification of malicious traffic patterns, enhancing overall network security.

DDoS attack detection and mitigation represents another crucial application area. Li et al.[74] developed DoLLM, which leverages LLaMA2-7B to understand non-contextual network flow data for detecting Carpet Bombing DDoS attacks. By reorganizing network flows into structured sequences and projecting them into the LLM's semantic space, DoLLM achieved F1 score improvements of up to 33.3% in zero-shot scenarios and at least 20.6% in real ISP traces compared to traditional methods. Similarly, Wang et al.[137] introduced ShieldGPT, an LLM-based framework that provides detailed explanatory analyses and effective mitigation strategies for DDoS attacks, demonstrating LLMs' ability to not only detect threats but also suggest actionable countermeasures.

Authentication security advances through LLM integration with blockchain technology. Pan et al.[105] combined BERT with blockchain for network identity authentication, achieving over 90% accuracy compared to 80-90% for traditional methods. This hybrid approach reduced replay attack success rates from 24.93% to 0.92% while maintaining low response times between 0.51-0.99 seconds. This integration demonstrates how LLMs can enhance security mechanisms while maintaining operational efficiency.

Protocol testing and vulnerability assessment benefit from LLMs' ability to understand complex specifications. Kakarla et al.[66] presented Eywa, which leverages GPT-4 to extract protocol specifications from RFCs and automatically generate test cases through symbolic execution. This approach revealed 26 unique bugs in ten DNS implementations, including 11 previously undiscovered bugs, demonstrating LLMs' effectiveness in identifying vulnerabilities that evaded detection by manual testing. Similarly, Meng et al.[95] developed CHATAFL, an LLM-guided protocol fuzzer that achieved 47.60% more state transition coverage and discovered nine previously unknown vulnerabilities in widely-used protocol implementations compared to state-of-the-art fuzzers.

Explainability in security decisions represents a distinctive advantage of LLM integration. Ali et al.[5] developed HuntGPT, which combines ML-based anomaly detection with GPT-3.5 Turbo to provide human-readable explanations of detected threats. Their system achieved cybersecurity knowledge success rates between 72-82.5% on standardized certification exams, demonstrating both technical accuracy and the ability to communicate complex security concepts effectively. Houssel et al.[55] explored using LLMs for explainable NIDS, finding that while LLMs may not surpass specialized models in pure detection tasks, they offer important value in explaining detected anomalies and aiding threat response when integrated with existing systems.

Intrusion prediction capabilities extend beyond detection to anticipate and mitigate threats proactively. Diaf et al.[35] developed a framework combining fine-tuned GPT-2 and BERT with LSTM

Table 8: Quantitative comparison of LLM-based approaches in network optimization

Topics	Baseline Method Performance	LLM Method Performance	Improvement
Knowledge-Free Resource Management [72]	Conventional optimization algorithms	Near-optimal energy efficiency after 100 iterations	Faster convergence with minimal domain knowledge
Multi-Expert Network Optimization [37]	Single DRL models for specific tasks	Above 85% task completion rate with MoE approach	Reduced need for task-specific models while maintaining performance
Game Theory-Based Optimization [50]	Expert-designed strategies	Matched expert performance while outperforming benchmarks	Automated strategy formulation without domain expertise
Graph-Based UAV Optimization [124]	Traditional GNN approaches	LLM+GNN outperformed LLM+GAT and LLM+Node2Vec	More stable and efficient dynamic network optimization
Multi-Objective SFC Optimization [76]	Model-based and greedy algorithms	Superior QoS with minimized payment costs	Balanced multiple competing objectives effectively
Network Slicing [12]	Traditional orchestration	500+ transactions per second for network slice trading	Scalable, transparent resource allocation
Task-Adaptive Networking [147]	Task-specific algorithms	10.1-36.6% error reduction; 14.5-36.6% QoE improvement	Single foundation model outperforming specialized solutions

Table 9: Papers on LLM Solutions for Network Optimization

Title	Ref	Year	Primary LLM
Generative AI for Game Theory-based Mobile Networking	[50]	2025	GPT-4
Large Language Model (LLM)-enabled Graphs in Dynamic Networking	[124]	2024	GPT-4 (ChatGPT), Bard, Bing Chat
Large Language Models for Knowledge-Free Network Management: Feasibility Study and Opportunities	[72]	2024	GPT-3.5-Turbo
Next-Gen Service Function Chain Deployment: Combining Multi-Objective Optimization with AI Large Language Models	[76]	2025	Moonshot-v1-32k
SliceGPT — OpenAI GPT-3.5 LLM, Blockchain and Non-Fungible Token Enabled Intelligent 5G/6G Network Slice Broker and Marketplace	[12]	2024	GPT-3.5-turbo
NetLLM: Adapting Large Language Models for Networking	[147]	2024	Llama2-7B
Mixture of Experts for Network Optimization: A Large Language Model-enabled Approach	[37]	2024	GPT-3.5-turbo
Reasoning AI Performance Degradation in 6G Networks with Large Language Models	[58]	2024	GPT-3.5, Multimodal-CoT-T5-223M, Multimodal-CoT-T5-738M
Reinforcement Learning-Based Load Balancing with Large Language Models and Edge Intelligence for Dynamic Cloud Environments	[32]	2023	GPT-4
Retrieval Augmented Generation with Multi-Modal LLM Framework for Wireless Environments	[97]	2025	GPT-4o, Gemini 1.5

networks to predict malicious activities in IoT networks, achieving 98% overall accuracy. Similarly, Adjewa et al.[3] created an adaptive framework using BERT embeddings with GMM for continuous detection of emerging attacks, maintaining 95.6% accuracy and recall even after integrating additional unknown attack clusters. These approaches demonstrate LLMs' ability to anticipate threats before they manifest, enabling proactive security measures.

Despite these advances, LLM applications in security face unique challenges. Houssel et al.[55] highlighted the 7000-fold slower inference time of LLMs compared to lightweight models, making them impractical for real-time NIDS applications without optimization. Additionally, privacy concerns remain paramount in security applications, driving innovations like Privacy-Preserving Fixed-Length Encoding [40] and locally-deployable models like Hackphyr [112], which achieved performance comparable to GPT-4 while running on a single GPU. Furthermore, Moskal et al.[99] explored ethical considerations regarding LLMs' potential to enhance threat actor capabilities, emphasizing the need for responsible deployment in security contexts.

As shown in Table 10, LLM-based approaches consistently outperform traditional security methods across various domains. These advancements demonstrate LLMs' potential to transform network

Table 10: Quantitative comparison of LLM-based approaches methods in network security

Security Approach	Baseline Method Performance	LLM Method Performance	Improvement
IoT Anomaly Detection [40]	Traditional ML and DL methods	98.2% accuracy with SecurityBERT	Higher accuracy with smaller model size (16.7MB)
Network Traffic Classification [154]	Naïve Bayes, SVM, MLP	94.84% accuracy with fine-tuned GPT-3.5	Significant performance improvement across diverse traffic types
DDoS Detection [74]	XGBoost, MLP, SVM	33.3% F1 score improvement in zero-shot scenarios	Better generalization to unseen attack patterns
Authentication Security [105]	Traditional methods	Blockchain + BERT Integration	Enhanced security with minimal latency impact
Protocol Testing [95]	State-of-the-art fuzzers (AFLNET, NSFUZZ)	47.60% more state transitions; 9 new vulnerabilities discovered	More comprehensive protocol testing with fewer resources
Continuous Threat Detection [3]	Static detection models	95.6% accuracy with novel attack integration	Adaptable detection without requiring retraining

security by enhancing detection accuracy, enabling proactive threat prevention, improving explainability, and facilitating more efficient vulnerability discovery—ultimately creating more robust and adaptable security solutions for modern network environments. Table 11 summarizes research papers leveraging LLMs for network security applications, demonstrating approaches for threat detection, explainability, and automated vulnerability assessment.

7 Challenges and Future Directions Of LLM-enabled NO&M

7.1 Challenges and Open Issues

Despite the advancements in integrating LLMs into NO&M, several technical challenges persist. These challenges stem from the inherent complexities of network environments, the limitations of current AI technologies, and the practical constraints of deploying LLMs in production settings. Based on our analysis of recent literature, we categorize these challenges into several key dimensions:

7.1.1 Scalability and Computational Efficiency. Current LLMs face scalability challenges in large-scale network deployments. As networks grow in complexity, the computational resources required for real-time processing and analysis become increasingly demanding. This challenge manifests in several ways:

Real-time Response Constraints: LLM inference latency ranges from 0.51 to 90 seconds according to empirical measurements in production networking systems [105, 148]. This latency presents a obstacle for latency-sensitive applications in network operations where sub-second responses are often necessary. As Houssel et al. [55] demonstrated, LLMs can be approximately 7000 times slower than lightweight models during inference, making them potentially unsuitable for time-critical network operations without optimization.

Resource-Constrained Environments: Modern LLMs typically require high-end GPUs with substantial memory, which conflicts with the limited computational resources and storage capacities available on many network devices. Shao et al. [120] identify this misalignment as a obstacle to widespread deployment in resource-constrained environments, particularly at network edges where compute resources are often limited.

Processing Telemetry Data: A technical challenge lies in adapting LLMs, which are primarily designed for natural language, to effectively process and analyze network telemetry data that is inherently structured and numerical. Huang et al. [58] propose embedding telemetry data into token-like representations, which enables LLMs to interpret heterogeneous network metrics more accurately. Other approaches proposed by Li et al. [74] include fine-tuning on structured datasets specifically for network operations, while Wu et al. [147] suggest integrating specialized pre-processing layers designed for networking data.

Recent research by Lee et al. [72] has explored knowledge-free optimization techniques whose operations are independent of scenario-specific information, achieving near-optimal energy efficiency within 100 iterations. However, these approaches still face challenges in balancing computational requirements with performance in resource-constrained environments.

7.1.2 Data Privacy and Security. The integration of LLMs into network operations introduces data privacy and security challenges:

Data Leakage Risks: LLMs require access to network configurations, traffic patterns, and other sensitive operational data, raising concerns about potential data breaches. This is particularly problematic when deploying commercial LLMs as services, where internal network information might leave organizational boundaries. Ferrag et al. [40] demonstrated these concerns by developing a Privacy-Preserving Fixed-Length Encoding technique specifically to address data leakage risks in security-sensitive contexts.

Training Data Contamination: The challenge of controlling what information LLMs absorb during training can lead to unintentional memorization of sensitive data. Ali and Kostakos [5] highlight the risk of LLMs potentially exposing proprietary network configurations or security vulnerabilities present in training data.

Compliance Challenges: LLMs deployed in network operations must adhere to industry regulations and compliance standards. As network management often involves regulated environments with strict compliance requirements, Rigaki et al. [112] developed Hackphyr, a locally fine-tuned LLM for network security environments that can operate independently of cloud services specifically to address privacy and compliance concerns.

Table 11: Papers on LLM Solutions for Network Security

Title	Ref	Year	Primary LLM
Anomaly Based Intrusion Detection using Large Language Models	[83]	2024	BERT
Combining BERT and Blockchain Technology to Improve Network Authentication Security	[105]	2024	BERT
Enhancing Network Traffic Classification with Large Language Models	[154]	2024	GPT-3.5-Turbo, GPT-4o
Hackphyr: A Local Fine-Tuned LLM Agent for Network Security Environments	[112]	2024	Zephyr-7B-β
Oracle-based Protocol Testing with Eywa	[66]	2023	GPT-4
ShieldGPT: An LLM-based Framework for DDoS Mitigation	[137]	2024	GPT-4
LLMs Killed the Script Kiddie: How Agents Supported by Large Language Models Change the Landscape of Network Threat Testing	[99]	2023	GPT-3.5-Turbo
Revolutionizing Cyber Threat Detection With Large Language Models: A Privacy-Preserving BERT-Based Lightweight Model for IoT/IIoT Devices	[40]	2024	BERT
Towards Explainable Network Intrusion Detection using Large Language Models	[55]	2024	GPT-4, Llama3-8B-Instruct
HuntGPT: Integrating Machine Learning-Based Anomaly Detection and Explainable AI with Large Language Models (LLMs)	[5]	2023	GPT-3.5 Turbo
Maximizing Penetration Testing Success with Effective Reconnaissance Techniques using ChatGPT	[128]	2023	ChatGPT
DoLLM: How Large Language Models Understanding Network Flow Data to Detect Carpet Bombing DDoS	[74]	2024	LLaMA2-7B
Large Language Model guided Protocol Fuzzing	[95]	2024	GPT-3.5 Turbo
Beyond Detection: Leveraging Large Language Models for Cyber Attack Prediction in IoT Networks	[35]	2024	GPT-2, BERT
GPT-2C: A Parser for Honeypot Logs Using Large Pre-trained Language Models	[118]	2021	GPT-2
LLM-based Continuous Intrusion Detection Framework for Next-Gen Networks	[3]	2024	BERT
Retrieval Augmented Generation Based LLM Evaluation For Protocol State Machine Inference With Chain-of-Thought Reasoning	[87]	2025	Llama-3-8B, Gemma-2-9B, text-embedding-ada-002

Recent approaches to address these issues include federated learning implementations [72], homomorphic encryption integration [40], and the development of smaller, locally deployable models like Hackphyr [112] that achieved performance comparable to much larger commercial models such as GPT-4 while running on a single GPU.

7.1.3 Explainability and Transparency. The black-box nature of LLMs presents challenges for their adoption in network operations:

Decision Traceability: Understanding how and why LLMs arrive at specific network management decisions is crucial for trust and accountability. Ali and Kostakos [5] highlight that network administrators require transparency in automated decision-making, especially for operations like security incident response or major configuration changes.

Causal Reasoning Limitations: Current LLMs struggle with establishing clear causal relationships in complex network environments. Houssel et al. [55] demonstrated that while LLMs can generate plausible explanations for network anomalies, they often lack the precision and causal reasoning capabilities required for debugging tasks, particularly in complex multi-domain networks.

Balancing Explainability with Performance: There exists a tension between model performance and explainability. Research by Ali and Kostakos [5] with HuntGPT demonstrated that incorporating explicit explanation mechanisms such as SHAP and Lime can enhance user understanding but may increase processing overhead and complexity.

Recent innovations addressing these challenges include attention visualization techniques [55], counterfactual explanations [5], and the development of hybrid models integrating traditional rule-based systems with LLMs. HuntGPT [5] successfully demonstrated knowledge success rates between 72-82.5% on standardized certification exams while maintaining human-readable explanations, showing that explainability and performance can be balanced with the right architectural approach.

7.1.4 Integration with Legacy Systems and Interoperability. The heterogeneous nature of network environments creates integration challenges:

Cross-Platform Configuration Translation: Network environments typically consist of devices from multiple vendors with different configuration syntaxes. Wei et al. [143] developed an IRAG framework that achieved 97.74% syntax correctness, outperforming GPT-4o by 19.07% in translating configurations across vendor-specific systems.

API Standardization Challenges: The lack of standardized APIs for LLM integration with existing network management systems creates implementation complexity. Many legacy systems were not designed with AI integration capabilities, requiring complex middleware or custom integration layers.

Automating Low-Level Network Configurations: Traditional approaches to network automation often rely on rigid rule-based scripting, which does not scale well to dynamic environments. Zhou et al. [152] demonstrated that NDTs can provide a novel approach to this challenge, enabling LLMs to generate structured, compliant, and automated network configurations while adapting to evolving policies.

Recent advancements include the development of open middleware platforms for LLM integration as shown in Lin et al.'s AppleSeed system [77], which achieved efficient intent translation with an average 22.3x lines of code to intent word ratio, while speeding up management plan execution by 1.7-2.6 times through just-in-time compilation when compared to sequential implementations.

7.1.5 Adaptation to Dynamic Network Environments. The rapidly evolving nature of network environments creates challenges for LLMs:

Handling Dynamic Network Topologies: Networks frequently undergo changes in topology, traffic patterns, and connectivity, requiring models to adapt without complete retraining. Shokrnezhad and Taleb [123] proposed an Autonomous Reinforcement Coordination framework that combines LLMs with hierarchical action planning and reinforcement learning to adapt to dynamic SAGINs environments.

Real-Time Decision Making: Network operations often require immediate responses to changing conditions. The inherent latency in LLM processing presents challenges for time-sensitive network operations. Manias et al. [90] addressed this by developing semantic routing architectures specifically designed to improve LLM performance in 5G core networks, demonstrating enhanced accuracy and efficiency compared to standalone LLM prompting methods.

Continuous Learning Mechanisms: As network environments evolve, LLMs need mechanisms to incorporate new knowledge without catastrophic forgetting. Adjewa et al. [3] developed a framework for continuous intrusion detection using BERT embeddings with Gaussian Mixture Models (GMM), maintaining 95.6% accuracy and recall even after integrating additional unknown attack clusters.

Recent research has shown promise in addressing these challenges through approaches like reinforcement learning-based adaptation [123], hybrid architectures combining LLMs with traditional machine learning [101], and continual learning mechanisms [3]. These approaches demonstrate the potential for LLMs to maintain effectiveness even in highly dynamic network environments.

7.1.6 Energy Efficiency and Sustainability. The computational intensity of LLMs raises concerns about their environmental impact and operational costs:

Training Energy Requirements: Training state-of-the-art LLMs can require substantial computational resources and energy, with environmental impacts that cannot be ignored in large-scale deployments. Zhang et al. [148] highlight that this often conflicts with organizational sustainability goals.

Inference Power Consumption: Even in inference mode, large models like Llama 70B require significant power, as demonstrated by Mekrache et al. [94], who noted that their best-performing open-source LLM for zero-touch network management required substantial energy resources.

Balancing Model Size and Performance: Finding the optimal balance between model capabilities and energy efficiency remains a challenge. Mekrache et al. [94] advocate for developing Small Language Models with fewer parameters (e.g., 1B) that can approach the efficiency of larger counterparts while consuming less power.

Recent innovations in addressing these challenges include model distillation techniques, quantization approaches that reduce computational requirements [148], and the development of specialized, domain-specific smaller models that offer comparable performance for networking tasks with reduced energy footprints [94].

7.1.7 Domain Knowledge Acquisition and Validation. Many current LLMs lack specialized domain knowledge required for effective NO&M:

Limited Networking Expertise: General-purpose LLMs trained on broad internet data often lack specialized knowledge of networking protocols, equipment, and operations. Shao et al. [120] identified this as a challenge that prevents LLMs from effectively addressing domain-specific networking problems.

Validating Generated Configurations: Network configurations generated by LLMs require careful validation to avoid potentially catastrophic errors. Mondal et al. [98] demonstrated that even advanced models like GPT-4 can produce configurations with egregious errors in topology, syntax, and semantics if used without appropriate verification.

Keeping Pace with Networking Evolution: The networking field evolves rapidly, presenting challenges for LLMs to remain current with the latest protocols, technologies, and best practices. Bansal et al. [13] demonstrated this challenge when applying LLMs to FCC compliance tasks, where accuracy of off-the-shelf models like GPT-4 (73.20%), ChatGPT (46.42%), and LLaMA (35.71%) was insufficient without domain-specific adaptation.

Recent approaches to address these challenges include Verified Prompt Programming [98] which achieved a 10× leverage ratio for Juniper translations and 6× for no-transit policies, Retrieval Augmented Generation [13, 143] which achieved a 51.8% improvement over baseline LLM performance, and domain-specific fine-tuning as demonstrated by Kan et al.'s Mobile-LLaMA [67], which outperformed GPT-3.5 in IP routing analysis.

7.2 Future Perspective and Research Directions

Based on our analysis of the current state of LLMs in NO&M and the challenges identified, we outline several promising research directions that could advance this field:

7.2.1 Developing Domain-Specific Efficient LLM Architectures. Future research should focus on creating more efficient LLM architectures specifically designed for network operations:

Network-Optimized Model Architectures: Rather than relying on general-purpose LLMs, specialized architectures customized for network operations could improve performance and efficiency. Kan et al. [67] demonstrated this approach with Mobile-LLaMA, an instruction-finetuned variant of LLaMA 2 13B specifically designed for network analysis in 5G environments.

Parameter-Efficient Fine-Tuning: Techniques like LoRA have shown promise in adapting LLMs to networking tasks with minimal additional parameters. Wu et al.'s NetLLM [147] leveraged LoRA-based fine-tuning to adapt Llama2-7B for networking tasks, reducing mean absolute error by 10.1-36.6% for viewport prediction while minimizing additional parameters.

Hardware-Aware Model Optimization: Future research should explore model designs that are optimized for the specific hardware constraints of network devices. Mekrache et al. [94] propose developing Green LLMs specifically optimized for deployment in networking environments with reduced power requirements.

The potential impact of these developments is big—they could enable the deployment of powerful LLM capabilities across a wider range of network environments, including resource-constrained edge devices, while reducing operational costs and environmental impact.

7.2.2 Advancing Intent-Based Network Management. IBN represents one of the promising applications of LLMs in network operations:

Multimodal Intent Recognition: Future IBN systems should integrate text, visual, and structured data inputs to better understand and implement network intents. Brodimas et al. [21] proposed leveraging multimodal generative AI to establish translation pipelines for network configuration that bridge user-friendly intent expression with underlying traditional network domains.

Automated Intent Validation: Research should focus on developing more robust techniques for validating that network configurations correctly implement user intents. Mekrache et al. [93] identified this as a component of Intent Life-Cycle management, proposing an LLM-centric Intent architecture that achieved initial rating scores of approximately 4.5 for intent decomposition.

Closed-Loop Intent Assurance: Future systems should autonomously monitor network performance to ensure intents are continuously satisfied, even as network conditions change. Asif et al. [11] demonstrated the potential of combining LLMs with KNN classification to achieve up to 5% higher accuracy (88%) in intent validation compared to existing methods.

These advancements could transform how networks are managed, making them more autonomous, adaptive, and aligned with business objectives rather than technical specifications.

7.2.3 Enhancing Security and Privacy Preservation. As networks face increasingly sophisticated threats, LLMs offer promising capabilities for security enhancement:

Explainable Network Security: Future research should focus on making LLM-based security decisions fully transparent and explainable. Ali and Kostakos [5] developed HuntGPT, which combined ML-based anomaly detection with GPT-3.5 Turbo to provide human-readable explanations of detected threats, achieving cybersecurity knowledge success rates between 72-82.5% on standardized certification exams.

Privacy-Preserving LLM Techniques: Research should explore techniques that allow LLMs to operate effectively on network data without compromising privacy. Ferrag et al. [40] introduced SecurityBERT, combining BERT with a novel Privacy-Preserving Fixed-Length Encoding technique to achieve 98.2% accuracy in classifying fourteen distinct attack types while maintaining privacy.

Proactive Threat Prediction: Rather than merely detecting known threats, LLMs could enable the prediction of potential threats before they materialize. Diaf et al. [35] demonstrated this possibility by combining fine-tuned GPT-2 and BERT with LSTM networks to predict malicious activities in IoT networks with 98% overall accuracy.

These developments could enhance network security postures while addressing the privacy concerns that currently limit the deployment of LLMs in sensitive network environments.

7.2.4 Integrating LLMs with Next-Generation Network Technologies. LLMs should be tightly integrated with network technologies to maximize their impact:

LLMs for 5G/6G Networks: Research should explore how LLMs can be leveraged to optimize resource allocation, spectrum management, and network slicing in next-generation wireless networks. Bandara et al. [12] developed SliceGPT, which employs a fine-tuned GPT-3.5-turbo model with blockchain and smart contracts to optimize network slice orchestration, supporting over 500 transactions per second.

Edge AI Integration: Future work should focus on deploying LLMs at the network edge to enable more responsive and efficient network operations. Shao et al. [120] identified this as a direction for enabling LLMs to operate effectively in resource-constrained environments with reduced latency.

NDT: LLMs can play a crucial role in creating and managing digital twins of network infrastructure. Zhou et al. [152] introduced Hermes, a framework that uses LLM agents to construct NDT instances through structured blueprints, achieving up to 80% accuracy in network modeling tasks.

These integrations could enable higher levels of network optimization, management, and intelligence across a wide range of network technologies.

7.2.5 Developing Green and Sustainable LLM Solutions. Given the environmental impact of LLMs, research into more sustainable approaches is essential:

Energy-Aware Model Selection: Future systems should automatically select the appropriate model size based on the specific network task requirements and available energy resources. Du et al. [37] proposed a MoE framework that dynamically selects specialized models based on user requirements, maintaining task

completion rates above 85% while optimizing computational efficiency.

Renewable Energy Integration: Research should explore how LLM-based network operations can be integrated with renewable energy sources to minimize environmental impact. This could involve scheduling computationally intensive operations during periods of renewable energy availability.

Model Compression and Distillation: Techniques that reduce model size while preserving performance are crucial for sustainable deployment. Mekrache et al. [94] advocated for developing Small Language Models with fewer parameters (e.g., 1B) as alternatives to power-hungry larger models.

These advancements could ensure that the benefits of LLMs in network operations are realized without unacceptable environmental consequences or operational costs.

7.2.6 Enabling Cross-Domain Collaboration. The complex, interdependent nature of modern networks requires collaborative approaches:

Multi-Agent LLM Systems: Research should explore how multiple specialized LLM agents can collaborate to manage different aspects of network operations. Shokrnezhad and Taleb [123] demonstrated the potential of this approach with their Autonomous Reinforcement Coordination framework combining LLM-based RAG with Hierarchical Action Planning.

Human-AI Collaborative Interfaces: Future systems should enable seamless collaboration between network operators and LLMs. Mondal et al.'s [98] Verified Prompt Programming demonstrated this approach by combining GPT-4 with verifiers and providing localized feedback to human operators.

Federated LLM Learning: Techniques that enable multiple organizations to train LLMs collaboratively without sharing sensitive network data could accelerate advancement. Lee et al. [72] identified federated techniques as promising approaches for enabling decentralized LLM optimizers distributed over cloud, edge, and devices.

These collaborative approaches could address the challenges of complex, heterogeneous network environments more effectively than isolated solutions.

7.2.7 Ethical and Responsible LLM Deployment. As LLMs become more integrated into network infrastructure, ethical considerations become increasingly important:

Fairness and Bias Mitigation: Research should address potential biases in LLM-based network management decisions. Moskal et al. [99] identified this as a concern, particularly when LLMs are used for security-related decisions that could disproportionately impact certain users or applications.

Accountability Frameworks: Clear mechanisms for establishing accountability in LLM-driven network operations are essential. Ali and Kostakos [5] emphasized the importance of this for maintaining trust in automated network operations, particularly when critical services are involved.

Transparent Decision Boundaries: Future systems should clearly delineate which decisions are made autonomously versus those requiring human approval. Lira et al. [78] proposed this approach for their LLM-NetCFG system, where certain configuration changes required explicit human verification.

By addressing these ethical considerations proactively, the field can avoid potential pitfalls and ensure that LLM deployment in network operations serves the broader public interest.

In summary, while LLMs offer capabilities for enhancing NO&M, realizing their full potential requires addressing technical, operational, and ethical challenges. The research directions outlined above represent promising pathways toward more intelligent, efficient, and sustainable network operations in the future. By pursuing these directions, the field can progress toward truly autonomous, adaptive, and resilient network management systems powered by advanced LLM capabilities.

8 Conclusion

This survey has provided an analysis of the application of LLMs in NO&M. We reviewed the advancements in integrating LLMs across various domains of networking, including Network Design, Network Automation, Network Optimization, and Network Security. The study highlighted the unique capabilities of LLMs in transforming traditional networking processes into more intelligent, autonomous, and efficient systems.

We systematically explored the challenges addressed by LLM-enabled techniques and discussed the solutions presented in the current literature. Our analysis has shed light on how LLMs can enhance network performance, enable intent-based management, improve security measures, and optimize resource allocation. We also identified the design aspects, including architecture, training methods, and integration strategies, that contribute to the effectiveness of LLMs in handling complex networking tasks.

Moreover, we emphasized the ongoing challenges such as scalability, data privacy, explainability, and the integration of LLMs with legacy systems, which still need to be addressed to realize the full potential of LLMs in network operations. Our Future Directions section outlined a detailed roadmap for future research, emphasizing the need for scalable architectures, energy-efficient models, improved security protocols, and the ethical use of AI in network environments.

This survey not only provides a detailed overview of the state-of-the-art research on LLMs in computer networking but also offers practical insights and future perspectives for researchers and practitioners. We anticipate that our findings will serve as a valuable resource for driving further innovation in the field, guiding the development of more advanced LLM-driven solutions that can meet the demands of next-generation network infrastructures.

References

[1] Asmaa Abdallah, Abdullatif Albaseer, Abdulkadir Celik, Mohamed Abdallah, and Ahmed M Eltawil. 2024. NetOrchLLM: Mastering Wireless Network Orchestration with Large Language Models. *arXiv preprint arXiv:2412.10107* (2024).

[2] Alibaba DAMO Academy. 2024. Qwen. https://github.com/QwenLM/Qwen Accessed: 2024-11-14.

[3] Frederic Adjewa, Moez Esseghir, and Leila Merghem-Boulahia. 2024. LLM-based Continuous Intrusion Detection Framework for Next-Gen Networks. *arXiv preprint arXiv:2411.03354* (2024).

[4] Abubakar S Ali, Dimitrios Michael Manias, Abdallah Shami, and Sami Muhaidat. 2023. Leveraging Large Language Models for DRL-Based Anti-Jamming Strategies in Zero Touch Networks. *arXiv preprint arXiv:2308.09376* (2023).

[5] Tarek Ali and Panos Kostakos. 2023. HuntGPT: Integrating machine learning-based anomaly detection and explainable AI with large language models (LLMs). *arXiv preprint arXiv:2309.16021* (2023).

[6] Ebtesam Almazrouei, Hamza Alobeidli, Abdulaziz Alshamsi, Alessandro Cappelli, Ruxandra Cojocaru, Mérouane Debbah, Étienne Goffinet, Daniel Hesslow,

Julien Launay, Quentin Malartic, et al. 2023. The falcon series of open language models. *arXiv preprint arXiv:2311.16867* (2023).

[7] Eric Anderson, Jonathan Fritz, Austin Lee, Bohou Li, Mark Lindblad, Henry Lindeman, Alex Meyer, Parth Parmar, Tanvi Ranade, Mehul A Shah, et al. 2024. The Design of an LLM-powered Unstructured Analytics System. *arXiv preprint arXiv:2409.00847* (2024).

[8] Anthropic. 2024. Claude: A language model designed for safety and helpfulness. https://www.anthropic.com/claude Accessed: 2024-11-14.

[9] Anthropic. 2024. Claude AI. https://claude.ai/ Accessed: 2024-11-13.

[10] Abi Aryan, Aakash Kumar Nain, Andrew McMahon, Lucas Augusto Meyer, and Harpreet Singh Sahota. 2023. The costly dilemma: generalization, evaluation and cost-optimal deployment of large language models. *arXiv preprint arXiv:2308.08061* (2023).

[11] Muhammad Asif, Talha Ahmed Khan, and Wang-Cheol Song. 2025. Evaluating Large Language Models for Optimized Intent Translation and Contradiction Detection using KNN in IBN. *IEEE Access* (2025).

[12] Eranga Bandara, Peter Foytik, Sachin Shetty, Ravi Mukkamala, Abdul Rahman, Xueping Liang, Ng Wee Keong, and Kasun De Zoysa. 2024. SliceGPT–OpenAI GPT-3.5 LLM, Blockchain and Non-Fungible Token Enabled Intelligent 5G/6G Network Slice Broker and Marketplace. In *2024 IEEE 21st Consumer Communications & Networking Conference (CCNC)*. IEEE, 439–445.

[13] Atul Bansal, Veronica Muriga, Jason Li, Lucy Duan, and Swarun Kumar. 2025. Can we make FCC Experts out of LLMs?. In *Proceedings of the 26th International Workshop on Mobile Computing Systems and Applications*. 85–90.

[14] Lina Bariah, Qiyang Zhao, Hang Zou, Yu Tian, Faouzi Bader, and Merouane Debbah. 2024. Large generative ai models for telecom: The next big thing? *IEEE Communications Magazine* 62, 11 (2024), 84–90.

[15] Batfish Project. 2024. Batfish: Network configuration analysis and validation. https://batfish.org/ Accessed: 2024-10-15.

[16] Emily M Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the dangers of stochastic parrots: Can language models be too big?. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*. 610–623.

[17] Sarthak Bhardwaj, Pardeep Singh, and Mohammad Khalid Pandit. 2024. A survey on the integration and optimization of large language models in edge computing environments. In *2024 16th International Conference on Computer and Automation Engineering (ICCAE)*. IEEE, 168–172.

[18] Dileesh Chandra Bikkasani and Malleswar Reddy Yerabolu. 2024. AI-Driven 5G Network Optimization: A Comprehensive Review of Resource Allocation, Traffic Management, and Dynamic Network Slicing. (2024).

[19] Andrei-Laurentiu Bornea, Fadhel Ayed, Antonio De Domenico, Nicola Piovesan, and Ali Maatouk. 2024. Telco-RAG: Navigating the challenges of retrieval-augmented language models for telecommunications. *arXiv preprint arXiv:2404.15939* (2024).

[20] Giampaolo Bovenzi, Francesco Cerasuolo, Domenico Ciuonzo, Davide Di Monda, Idio Guarino, Antonio Montieri, Valerio Persico, and Antonio Pescapè. 2025. Mapping the Landscape of Generative AI in Network Monitoring and Management. arXiv:2502.08576 [cs.NI] https://arxiv.org/abs/2502.08576

[21] Dimitrios Brodimas, Kostis Trantzas, Besiana Agko, Georgios Christos Tziavas, Christos Tranoris, Spyros Denazis, and Alexios Birbas. 2024. Towards Intent-based Network Management for the 6G System adopting Multimodal Generative AI. In *2024 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit)*. IEEE, 848–853.

[22] Tom B Brown. 2020. Language models are few-shot learners. *arXiv preprint arXiv:2005.14165* (2020).

[23] Abdulkadir Celik and Ahmed M Eltawil. 2024. At the Dawn of Generative AI Era: A tutorial-cum-survey on new frontiers in 6G wireless intelligence. *IEEE Open Journal of the Communications Society* (2024).

[24] Supratim Chakraborty, Nithin Chitta, and Rajesh Sundaresan. 2024. Automation of Network Configuration Generation using Large Language Models. In *2024 20th International Conference on Network and Service Management (CNSM)*. IFIP, 1–9.

[25] Abdelaali Chaoub and Muslim Elkotob. 2025. Mobile Network-specialized Large Language Models for 6G: Architectures, Innovations, Challenges, and Future Trends. arXiv:2502.04933 [cs.NI] https://arxiv.org/abs/2502.04933

[26] Handi Chen, Weipeng Deng, Shuo Yang, Jinfeng Xu, Zhihan Jiang, Edith CH Ngai, Jiangchuan Liu, and Xue Liu. 2024. Towards Edge General Intelligence via Large Language Models: Opportunities and Challenges. *arXiv preprint arXiv:2410.18125* (2024).

[27] Shouyuan Chen, Sherman Wong, Liangjian Chen, and Yuandong Tian. 2023. Extending context window of large language models via positional interpolation. *arXiv preprint arXiv:2306.15595* (2023).

[28] Cisco Press. 2011. Cisco's PPDIOO Network Cycle. https://www.ciscopress.com/articles/article.asp?p=1697888 [Online; accessed 4-April-2025].

[29] Alexander Clemm, Laurent Ciavaglia, Lisandro Zambenedetti Granville, and Jeff Tantsura. 2022. Intent-Based Networking - Concepts and Definitions. RFC 9315. https://doi.org/10.17487/RFC9315

[30] Jan Clusmann, Fiona R Kolbinger, Hannah Sophie Muti, Zunamys I Carrero, Jan-Niklas Eckardt, Narmin Ghaffari Laleh, Chiara Maria Lavinia Löffler, Sophie-Caroline Schwarzkopf, Michaela Unger, Gregory P Veldhuizen, et al. 2023. The future landscape of large language models in medicine. *Communications medicine* 3, 1 (2023), 141.

[31] Abdulhalim Dandoush, Viswanath Kumarskandpriya, Mueen Uddin, and Usman Khalil. 2024. Large Language Models meet Network Slicing Management and Orchestration. *arXiv preprint arXiv:2403.13721* (2024).

[32] Bhavin Desai and Kapil Patil. 2023. Reinforcement learning-based load balancing with large language models and edge intelligence for dynamic cloud environments. *Journal of Innovative Technologies* 6, 1 (2023), 1–13.

[33] Bhavin Desai, Kapil Patil, Asit Patil, and Ishita Mehta. 2023. Large Language Models: A Comprehensive Exploration of Modern AI's Potential and Pitfalls. *Journal of Innovative Technologies* 6, 1 (2023).

[34] Jacob Devlin. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).

[35] Alaeddine Diaf, Abdelaziz Amara Korba, Nour Elislem Karabadji, and Yacine Ghamri-Doudane. 2024. Beyond Detection: Leveraging Large Language Models for Cyber Attack Prediction in IoT Networks. In *2024 20th International Conference on Distributed Computing in Smart Systems and the Internet of Things (DCOSS-IoT)*. IEEE, 117–123.

[36] Ning Ding, Yujia Qin, Guang Yang, Fuchao Wei, Zonghan Yang, Yusheng Su, Shengding Hu, Yulin Chen, Chi-Min Chan, Weize Chen, et al. 2023. Parameter-efficient fine-tuning of large-scale pre-trained language models. *Nature Machine Intelligence* 5, 3 (2023), 220–235.

[37] Hongyang Du, Guangyuan Liu, Yijing Lin, Dusit Niyato, Jiawen Kang, Zehui Xiong, and Dong In Kim. 2024. Mixture of Experts for Network Optimization: A Large Language Model-enabled Approach. *arXiv preprint arXiv:2402.09756* (2024).

[38] Kristina Dzeparoska, Jieyu Lin, Ali Tizghadam, and Alberto Leon-Garcia. 2023. LLM-based policy generation for intent-based management of applications. In *2023 19th International Conference on Network and Service Management (CNSM)*. IEEE, 1–7.

[39] Yifei Fang, Ke Lu, Jian Xue, Fuqing Li, and Zhentong Lyu. 2024. LLMNDC: A Novel Approach for Network Device Configuration based on Fine-tuned Large Language Models. In *2024 5th International Conference on Computer Engineering and Intelligent Control (ICCEIC)*. IEEE, 283–289.

[40] Mohamed Amine Ferrag, Mthandazo Ndhlovu, Norbert Tihanyi, Lucas C Cordeiro, Merouane Debbah, Thierry Lestable, and Narinderjit Singh Thandi. 2024. Revolutionizing cyber threat detection with large language models: A privacy-preserving bert-based lightweight model for iot/iiot devices. *IEEE Access* (2024).

[41] John Fields, Kevin Chovanec, and Praveen Madiraju. 2024. A survey of text classification with transformers: How wide? how large? how long? how accurate? how expensive? how safe? *IEEE Access* (2024).

[42] Ahlam Fuad, Azza H Ahmed, Michael A Riegler, and Tarik Čičić. 2024. An Intent-based Networks Framework based on Large Language Models. In *2024 IEEE 10th International Conference on Network Softwarization (NetSoft)*. IEEE, 7–12.

[43] Jorge Gallego-Madrid, Ramon Sanchez-Iborra, Pedro M Ruiz, and Antonio F Skarmeta. 2022. Machine learning-based zero-touch network and service management: A survey. *Digital Communications and Networks* 8, 2 (2022), 105–123.

[44] Leo Gao, Stella Biderman, Sid Black, Laurence Golding, Travis Hoppe, Charles Foster, Jason Phang, Horace He, Anish Thite, Noa Nabeshima, et al. 2020. The pile: An 800gb dataset of diverse text for language modeling. *arXiv preprint arXiv:2101.00027* (2020).

[45] Hamid Ghasemirahni, Alireza Farshin, Mariano Scazzariello, Marco Chiesa, and Dejan Kostić. 2024. Deploying Stateful Network Functions Efficiently using Large Language Models. In *Proceedings of the 4th Workshop on Machine Learning and Systems*. 28–38.

[46] Lingqi Guo, Jingyu Wang, Jianyu Wu, Caijun Yan, Haifeng Sun, Zirui Zhuang, Qi Qi, Yi Dong, Haibao Ren, and Jianxin Liao. 2024. Following the Compass: LLM-Empowered Intent Translation with Manual Guidance. In *2024 IEEE 32nd International Conference on Network Protocols (ICNP)*. IEEE, 1–12.

[47] Md Arafat Habib, Pedro Enrique Iturria Rivera, Yigit Ozcan, Medhat Elsayed, Majid Bavand, Raimundus Gaigalas, and Melike Erol-Kantarci. 2024. LLM-Based Intent Processing and Network Optimization Using Attention-Based Hierarchical Reinforcement Learning. *arXiv preprint arXiv:2406.06059* (2024).

[48] Ching-Nam Hang, Pei-Duo Yu, Roberto Morabito, and Chee-Wei Tan. 2024. Large Language Models Meet Next-Generation Networking Technologies: A Review. *Future Internet* 16, 10 (2024), 365.

[49] SA Hassan, MS Omar, MA Imran, J Qadir, and D Jayako. 2023. Performance Enhancement of 5G Networks Using AI-Driven Techniques. *Int. J. Appl. Res. Technol* 12 (2023), 64–70.

[50] Long He, Geng Sun, Dusit Niyato, Hongyang Du, Fang Mei, Jiawen Kang, Mérouane Debbah, and Zhu Han. 2025. Generative ai for game theory-based mobile networking. *IEEE Wireless Communications* 32, 1 (2025), 122–130.

[51] Zhiyuan He, Aashish Gottipati, Lili Qiu, Xufang Luo, Kenuo Xu, Yuqing Yang, and Francis Y Yan. 2024. Designing Network Algorithms via Large Language Models. In *Proceedings of the 23rd ACM Workshop on Hot Topics in Networks*. 205–212.

[52] Zhiyuan He, Aashish Gottipati, Lili Qiu, Francis Y Yan, Xufang Luo, Kenuo Xu, and Yuqing Yang. 2024. LLM-ABR: Designing Adaptive Bitrate Algorithms via Large Language Models. *arXiv preprint arXiv:2404.01617* (2024).

[53] Ekram Hossain and Monowar Hasan. 2015. 5G cellular: key enabling technologies and research challenges. *IEEE Instrumentation & Measurement Magazine* 18, 3 (2015), 11–21.

[54] Xinyi Hou, Yanjie Zhao, Yue Liu, Zhou Yang, Kailong Wang, Li Li, Xiapu Luo, David Lo, John Grundy, and Haoyu Wang. 2023. Large language models for software engineering: A systematic literature review. *ACM Transactions on Software Engineering and Methodology* (2023).

[55] Paul RB Houssel, Priyanka Singh, Siamak Layeghy, and Marius Portmann. 2024. Towards Explainable Network Intrusion Detection using Large Language Models. *arXiv preprint arXiv:2408.04342* (2024).

[56] Jeremy Howard and Sebastian Ruder. 2018. Universal language model fine-tuning for text classification. *arXiv preprint arXiv:1801.06146* (2018).

[57] Fei Hu, Qi Hao, and Ke Bao. 2014. A survey on software-defined network and openflow: From concept to implementation. *IEEE Communications Surveys & Tutorials* 16, 4 (2014), 2181–2206.

[58] Liming Huang, Yulei Wu, and Dimitra Simeonidou. 2024. Reasoning AI Performance Degradation in 6G Networks with Large Language Models. *arXiv preprint arXiv:2408.17097* (2024).

[59] Beni Ifland, Elad Duani, Rubin Krief, Miro Ohana, Aviram Zilberman, Andres Murillo, Ofir Manor, Ortal Lavi, Hikichi Kenji, Asaf Shabtai, et al. 2024. GeNet: A Multimodal LLM-Based Co-Pilot for Network Topology and Configuration. *arXiv preprint arXiv:2407.08249* (2024).

[60] Shumaila Javaid, Ruhul Amin Khalil, Nasir Saeed, Bin He, and Mohamed-Slim Alouini. 2024. Leveraging large language models for integrated satellite-aerial-terrestrial networks: recent advances and future directions. *arXiv preprint arXiv:2407.04581* (2024).

[61] Cheonsu Jeong. 2024. Fine-tuning and utilization methods of domain-specific llms. *arXiv preprint arXiv:2401.02981* (2024).

[62] Eui-Dong Jeong, Hee-Gon Kim, Sukhyun Nam, Jae-Hyoung Yoo, and James Won-Ki Hong. 2024. S-Witch: Switch Configuration Assistant with LLM and Prompt Engineering. In *NOMS 2024-2024 IEEE Network Operations and Management Symposium*. IEEE, 1–7.

[63] Albert Q Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, et al. 2023. Mistral 7B. *arXiv preprint arXiv:2310.06825* (2023).

[64] Albert Q Jiang, Alexandre Sablayrolles, Antoine Roux, Arthur Mensch, Blanche Savary, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Emma Bou Hanna, Florian Bressand, et al. 2024. Mixtral of experts. *arXiv preprint arXiv:2401.04088* (2024).

[65] Jean Kaddour, Joshua Harris, Maximilian Mozes, Herbie Bradley, Roberta Raileanu, and Robert McHardy. 2023. Challenges and applications of large language models. *arXiv preprint arXiv:2307.10169* (2023).

[66] Siva Kesava Reddy Kakarla and Ryan Beckett. 2023. Oracle-based Protocol Testing with Eywa. *arXiv preprint arXiv:2312.06875* (2023).

[67] Khen Bo Kan, Hyunsu Mun, Guohong Cao, and Youngseok Lee. 2024. Mobile-LLaMA: Instruction Fine-Tuning Open-Source LLM for Network Analysis in 5G Networks. *IEEE Network* (2024).

[68] Fahime Khoramnejad and Ekram Hossain. 2025. Generative AI for the optimization of next-generation wireless networks: Basics, state-of-the-art, and open challenges. *IEEE Communications Surveys & Tutorials* (2025).

[69] Hyojoon Kim and Nick Feamster. 2013. Improving network management with software defined networking. *IEEE Communications magazine* 51, 2 (2013), 114–119.

[70] Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. 2022. Large language models are zero-shot reasoners. *Advances in neural information processing systems* 35 (2022), 22199–22213.

[71] Taku Kudo. 2018. Subword regularization: Improving neural network translation models with multiple subword candidates. *arXiv preprint arXiv:1804.10959* (2018).

[72] Hoon Lee, Mintae Kim, Seunghwan Baek, Namyoon Lee, Merouane Debbah, and Inkyu Lee. 2024. Large Language Models for Knowledge-Free Network Management: Feasibility Study and Opportunities. *arXiv preprint arXiv:2410.17259* (2024).

[73] Fuliang Li, Haozhi Lang, Jiajie Zhang, Jiaxing Shen, and Xingwei Wang. 2024. PreConfig: A Pretrained Model for Automating Network Configuration. *arXiv preprint arXiv:2403.09369* (2024).

[74] Qingyang Li, Yihang Zhang, Zhidong Jia, Yannan Hu, Lei Zhang, Jianrong Zhang, Yongming Xu, Yong Cui, Zongming Guo, and Xinggong Zhang. 2024. DoLLM: How Large Language Models Understanding Network Flow Data to Detect Carpet Bombing DDoS. *arXiv preprint arXiv:2405.07638* (2024).

[75] Yuanzhi Li, Sébastien Bubeck, Ronen Eldan, Allie Del Giorno, Suriya Gunasekar, and Yin Tat Lee. 2023. Textbooks are all you need ii: phi-1.5 technical report. *arXiv preprint arXiv:2309.05463* (2023).

[76] Yuanfeng Li, Qi Zhang, Haipeng Yao, Ran Gao, Xiangjun Xin, and Mohsen Guizani. 2025. Next-Gen Service Function Chain Deployment: Combining Multi-Objective Optimization with AI Large Language Models. *IEEE Network* (2025).

[77] Jieyu Lin, Kristina Dzeparoska, Ali Tizghadam, and Alberto Leon-Garcia. 2023. Appleseed: Intent-based multi-domain infrastructure management via few-shot learning. In *2023 IEEE 9th International Conference on Network Softwarization (NetSoft)*. IEEE, 539–544.

[78] Oscar G Lira, Oscar M Caicedo, and Nelson LS da Fonseca. 2024. Large Language Models for Zero Touch Network Configuration Management. *arXiv preprint arXiv:2408.13298* (2024).

[79] Chang Liu, Xiaohui Xie, Xinggong Zhang, and Yong Cui. 2024. Large Language Models for Networking: Workflow, Advances and Challenges. *arXiv preprint arXiv:2404.12901* (2024).

[80] Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. 2023. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *Comput. Surveys* 55, 9 (2023), 1–35.

[81] Sasha Luccioni, Yacine Jernite, and Emma Strubell. 2024. Power hungry processing: Watts driving the cost of ai deployment?. In *Proceedings of the 2024 ACM conference on fairness, accountability, and transparency*. 85–99.

[82] Qing Lyu, Shreya Havaldar, Adam Stein, Li Zhang, Delip Rao, Eric Wong, Marianna Apidianaki, and Chris Callison-Burch. 2023. Faithful chain-of-thought reasoning. *arXiv preprint arXiv:2301.13379* (2023).

[83] Zineb Maasaoui, Mheni Merzouki, Abdella Battou, and Ahmed Lbath. 2024. Anomaly Based Intrusion Detection Using Large Language Models. In *2024 IEEE/ACS 21st International Conference on Computer Systems and Applications (AICCSA)*. IEEE, 1–8.

[84] Ali Maatouk, Kenny Chirino Ampudia, Rex Ying, and Leandros Tassiulas. 2024. Tele-LLMs: A Series of Specialized Large Language Models for Telecommunications. *arXiv preprint arXiv:2409.05314* (2024).

[85] Ali Maatouk, Fadhel Ayed, Nicola Piovesan, Antonio De Domenico, Merouane Debbah, and Zhi-Quan Luo. 2023. Teleqna: A benchmark dataset to assess large language models telecommunications knowledge. *arXiv preprint arXiv:2310.15051* (2023).

[86] Aman Madaan, Niket Tandon, Prakhar Gupta, Skyler Hallinan, Luyu Gao, Sarah Wiegreffe, Uri Alon, Nouha Dziri, Shrimai Prabhumoye, Yiming Yang, et al. 2024. Self-refine: Iterative refinement with self-feedback. *Advances in Neural Information Processing Systems* 36 (2024).

[87] Youssef Maklad, Fares Wael, Wael Elsersy, and Ali Hamdi. 2025. Retrieval Augmented Generation Based LLM Evaluation For Protocol State Machine Inference With Chain-of-Thought Reasoning. *arXiv preprint arXiv:2502.15727* (2025).

[88] Jiya Manchanda, Laura Boettcher, Matheus Westphalen, and Jasser Jasser. 2024. The Open Source Advantage in Large Language Models (LLMs). *arXiv preprint arXiv:2412.12004* (2024).

[89] Sathiya Kumaran Mani, Yajie Zhou, Kevin Hsieh, Santiago Segarra, Trevor Eberl, Eliran Azulai, Ido Frizler, Ranveer Chandra, and Srikanth Kandula. 2023. Enhancing network management using code generated by large language models. In *Proceedings of the 22nd ACM Workshop on Hot Topics in Networks*. 196–204.

[90] Dimitrios Michael Manias, Ali Chouman, and Abdallah Shami. 2024. Semantic Routing for Enhanced Performance of LLM-Assisted Intent-Based 5G Core Network Management and Orchestration. *arXiv preprint arXiv:2404.15869* (2024).

[91] Dimitrios Michael Manias, Ali Chouman, and Abdallah Shami. 2024. Towards Intent-Based Network Management: Large Language Models for Intent Extraction in 5G Core Networks. In *2024 20th International Conference on the Design of Reliable Communication Networks (DRCN)*. IEEE, 1–6.

[92] Abdelkader Mekrache and Adlen Ksentini. 2024. LLM-enabled Intent-driven Service Configuration for Next Generation Networks. In *2024 IEEE 10th International Conference on Network Softwarization (NetSoft)*. IEEE, 253–257.

[93] Abdelkader Mekrache, Adlen Ksentini, and Christos Verikoukis. 2024. Intent-based management of next-generation networks: an LLM-centric approach. *Ieee Network* (2024).

[94] Abdelkader Mekrache, Mohamed Mekki, Adlen Ksentini, Bouziane Brik, and Christos Verikoukis. 2024. On Combining XAI and LLMs for Trustworthy Zero-Touch Network and Service Management in 6G. *IEEE Communications Magazine* (2024).

[95] Ruijie Meng, Martin Mirchev, Marcel Böhme, and Abhik Roychoudhury. 2024. Large language model guided protocol fuzzing. In *Proceedings of the 31st Annual Network and Distributed System Security Symposium (NDSS)*.

[96] Yukai Miao, Yu Bai, Li Chen, Dan Li, Haifeng Sun, Xizheng Wang, Ziqiu Luo, Dapeng Sun, and Xiuting Xu. 2023. An empirical study of netops capability of pre-trained large language models. *arXiv preprint arXiv:2309.05557* (2023).

[97] Muhammad Ahmed Mohsin, Ahsan Bilal, Sagnik Bhattacharya, and John M Cioffi. 2025. Retrieval Augmented Generation with Multi-Modal LLM Framework for Wireless Environments. *arXiv preprint arXiv:2503.07670* (2025).

[98] Rajdeep Mondal, Alan Tang, Ryan Beckett, Todd Millstein, and George Varghese. 2023. What do llms need to synthesize correct router configurations?. In *Proceedings of the 22nd ACM Workshop on Hot Topics in Networks*. 189–195.

[99] Stephen Moskal, Sam Laney, Erik Hemberg, and Una-May O'Reilly. 2023. LLMs Killed the Script Kiddie: How Agents Supported by Large Language Models Change the Landscape of Network Threat Testing. *arXiv preprint arXiv:2310.06936* (2023).

[100] Uma Maheswari Natarajan, Raghuram Bharadwaj Diddigi, and Jyotsna Bapat. 2025. RAG-inspired Intent-Based Solution for Intelligent Autonomous Networks. In *2025 17th International Conference on COMmunication Systems and NETworks (COMSNETS)*. IEEE, 413–421.

[101] Hyeonho Noh, Byonghyo Shim, and Hyun Jong Yang. 2025. Adaptive Resource Allocation Optimization Using Large Language Models in Dynamic Wireless Environments. *arXiv preprint arXiv:2502.02287* (2025).

[102] ETSI Industry Specification Group (ISG) on Network Functions Virtualisation. 2012. *Network Functions Virtualisation: An Introduction, Benefits, Enablers, Challenges & Call for Action.* White Paper. European Telecommunications Standards Institute (ETSI). https://portal.etsi.org/nfv/nfv_white_paper.pdf

[103] OpenAI. 2024. ChatGPT. https://chatgpt.com/ Accessed: 2024-11-13.

[104] OpenAI. 2024. OpenAI Models overview. https://platform.openai.com/docs/models Accessed: 2024-10-15.

[105] Xiaojun Pan. 2024. Combining BERT and Blockchain Technology to Improve Network Authentication Security. In *2024 IEEE 7th International Conference on Automation, Electronics and Electrical Engineering (AUTEEE)*. IEEE, 883–888.

[106] Ivan Provilkov, Dmitrii Emelianenko, and Elena Voita. 2019. BPE-dropout: Simple and effective subword regularization. *arXiv preprint arXiv:1910.13267* (2019).

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

[108] Alec Radford. 2018. Improving language understanding by generative pre-training. (2018).

[109] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. 2019. Language models are unsupervised multitask learners. *OpenAI blog* 1, 8 (2019), 9.

[110] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of machine learning research* 21, 140 (2020), 1–67.

[111] Danish Rafique and Luis Velasco. 2018. Machine learning for network automation: overview, architecture, and applications [Invited Tutorial]. *Journal of Optical Communications and Networking* 10, 10 (2018), D126–D143.

[112] Maria Rigaki, Carlos Catania, and Sebastian Garcia. 2024. Hackphyr: A Local Fine-Tuned LLM Agent for Network Security Environments. *arXiv preprint arXiv:2409.11276* (2024).

[113] Baptiste Roziere, Jonas Gehring, Fabian Gloeckle, Sten Sootla, Itai Gat, Xiaoqing Ellen Tan, Yossi Adi, Jingyu Liu, Romain Sauvestre, Tal Remez, et al. 2023. Code llama: Open foundation models for code. *arXiv preprint arXiv:2308.12950* (2023).

[114] Walid S Saba. 2024. LLMs' Understanding of Natural Language Revealed. *arXiv preprint arXiv:2407.19630* (2024).

[115] Pranab Sahoo, Ayush Kumar Singh, Sriparna Saha, Vinija Jain, Samrat Mondal, and Aman Chadha. 2024. A systematic survey of prompt engineering in large language models: Techniques and applications. *arXiv preprint arXiv:2402.07927* (2024).

[116] Ahmed I Salameh and Mohamed El Tarhuni. 2022. From 5G to 6G—challenges, technologies, and applications. *Future Internet* 14, 4 (2022), 117.

[117] Rico Sennrich. 2015. Neural machine translation of rare words with subword units. *arXiv preprint arXiv:1508.07909* (2015).

[118] Febrian Setianto, Erion Tsani, Fatima Sadiq, Georgios Domalis, Dimitris Tsakalidis, and Panos Kostakos. 2021. GPT-2C: A parser for honeypot logs using large pre-trained language models. In *Proceedings of the 2021 IEEE/ACM international conference on advances in social networks analysis and mining*. 649–653.

[119] Amin Shahraki, Mahmoud Abbasi, Md Jalil Piran, and Amir Taherkordi. 2021. A comprehensive survey on 6G networks: Applications, core services, enabling technologies, and future challenges. *arXiv preprint arXiv:2101.12475* (2021).

[120] Jiawei Shao, Jingwen Tong, Qiong Wu, Wei Guo, Zijian Li, Zehong Lin, and Jun Zhang. 2024. WirelessLLM: Empowering Large Language Models Towards Wireless Intelligence. *arXiv preprint arXiv:2405.17053* (2024).

[121] Prakhar Sharma and Vinod Yegneswaran. 2023. PROSPER: Extracting Protocol Specifications Using Large Language Models. In *Proceedings of the 22nd ACM Workshop on Hot Topics in Networks*. 41–47.

[122] Xuemin Shen, Jie Gao, Wen Wu, Mushu Li, Conghao Zhou, and Weihua Zhuang. 2021. Holistic network virtualization and pervasive network intelligence for 6G. *IEEE Communications Surveys & Tutorials* 24, 1 (2021), 1–30.

[123] Masoud Shokrnezhad and Tarik Taleb. 2025. An Autonomous Network Orchestration Framework Integrating Large Language Models with Continual Reinforcement Learning. *arXiv preprint arXiv:2502.16198* (2025).

[124] Geng Sun, Yixian Wang, Dusit Niyato, Jiacheng Wang, Xinying Wang, H Vincent Poor, and Khaled B Letaief. 2024. Large language model (llm)-enabled graphs in dynamic networking. *IEEE Network* (2024).

[125] Tiao Tan, Fengxiao Tang, and Ming Zhao. 2025. Adapting Network Information to Semantics for Generalizable and Plug-and-Play Multi-Scenario Network Diagnosis. *arXiv preprint arXiv:2501.16842* (2025).

[126] Sasu Tarkoma, Roberto Morabito, and Jaakko Sauvola. 2023. AI-native interconnect framework for integration of large language model technologies in 6G systems. *arXiv preprint arXiv:2311.05842* (2023).

[127] Gemini Team, Rohan Anil, Sebastian Borgeaud, Jean-Baptiste Alayrac, Jiahui Yu, Radu Soricut, Johan Schalkwyk, Andrew M Dai, Anja Hauth, Katie Millican, et al. 2023. Gemini: a family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805* (2023).

[128] Sheetal Temara. 2023. Maximizing penetration testing success with effective reconnaissance techniques using chatgpt. *arXiv preprint arXiv:2307.06391* (2023).

[129] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. 2023. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971* (2023).

[130] A Vaswani. 2017. Attention is all you need. *Advances in Neural Information Processing Systems* (2017).

[131] Thai-Hoc Vu, Senthil Kumar Jagatheesaperumal, Minh-Duong Nguyen, Nguyen Van Huynh, Sunghwan Kim, and Quoc-Viet Pham. 2024. Applications of Generative AI (GAI) for Mobile and Wireless Networking: A Survey. *arXiv preprint arXiv:2405.20024* (2024).

[132] Changjie Wang, Mariano Scazzariello, Alireza Farshin, Simone Ferlin, Dejan Kostić, and Marco Chiesa. 2024. NetConfEval: Can LLMs Facilitate Network Configuration? *Proceedings of the ACM on Networking* 2, CoNEXT2 (2024), 1–25.

[133] Changjie Wang, Mariano Scazzariello, Alireza Farshin, Dejan Kostic, and Marco Chiesa. 2023. Making Network Configuration Human Friendly. *arXiv preprint arXiv:2309.06342* (2023).

[134] Dixuan Wang, Yanda Li, Junyuan Jiang, Zepeng Ding, Guochao Jiang, Jiaqing Liang, and Deqing Yang. 2024. Tokenization Matters! Degrading Large Language Models through Challenging Their Tokenization. *arXiv preprint arXiv:2405.17067* (2024).

[135] Jingyu Wang, Lei Zhang, Yiran Yang, Zirui Zhuang, Qi Qi, Haifeng Sun, Lu Lu, Junlan Feng, and Jianxin Liao. 2023. Network meets chatgpt: Intent autonomous management, control and operation. *Journal of Communications and Information Networks* 8, 3 (2023), 239–255.

[136] Mowei Wang, Yong Cui, Xin Wang, Shihan Xiao, and Junchen Jiang. 2017. Machine learning for networking: Workflow, advances and opportunities. *Ieee Network* 32, 2 (2017), 92–99.

[137] Tongze Wang, Xiaohui Xie, Lei Zhang, Chuyi Wang, Liang Zhang, and Yong Cui. 2024. ShieldGPT: An LLM-based Framework for DDoS Mitigation. In *Proceedings of the 8th Asia-Pacific Workshop on Networking*. 108–114.

[138] Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc Le, Ed Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou. 2022. Self-consistency improves chain of thought reasoning in language models. *arXiv preprint arXiv:2203.11171* (2022).

[139] Yaqing Wang, Quanming Yao, James T Kwok, and Lionel M Ni. 2020. Generalizing from a few examples: A survey on few-shot learning. *ACM computing surveys (csur)* 53, 3 (2020), 1–34.

[140] Jason Wei, Maarten Bosma, Vincent Y Zhao, Kelvin Guu, Adams Wei Yu, Brian Lester, Nan Du, Andrew M Dai, and Quoc V Le. 2021. Finetuned language models are zero-shot learners. *arXiv preprint arXiv:2109.01652* (2021).

[141] Jason Wei, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, Sebastian Borgeaud, Dani Yogatama, Maarten Bosma, Denny Zhou, Donald Metzler, et al. 2022. Emergent abilities of large language models. *arXiv preprint arXiv:2206.07682* (2022).

[142] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems* 35 (2022), 24824–24837.

[143] Yunze Wei, Xiaohui Xie, Yiwei Zuo, Tianshuo Hu, Xinyi Chen, Kaiwen Chi, and Yong Cui. 2025. Leveraging LLM Agents for Translating Network Configurations. *arXiv preprint arXiv:2501.08760* (2025).

[144] Jules White, Quchen Fu, Sam Hays, Michael Sandborn, Carlos Olea, Henry Gilbert, Ashraf Elnashar, Jesse Spencer-Smith, and Douglas C Schmidt. 2023. A prompt pattern catalog to enhance prompt engineering with chatgpt. *arXiv preprint arXiv:2302.11382* (2023).

[145] Wikipedia contributors. 2025. FCAPS — Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/wiki/FCAPS [Online; accessed 4-April-2025].

[146] Carole-Jean Wu, Ramya Raghavendra, Udit Gupta, Bilge Acun, Newsha Ardalani, Kiwan Maeng, Gloria Chang, Fiona Aga, Jinshi Huang, Charles Bai, et al. 2022.

Sustainable ai: Environmental implications, challenges and opportunities. *Proceedings of Machine Learning and Systems* 4 (2022), 795–813.

[147] Duo Wu, Xianda Wang, Yaqi Qiao, Zhi Wang, Junchen Jiang, Shuguang Cui, and Fangxin Wang. 2024. NetLLM: Adapting large language models for networking. In *Proceedings of the ACM SIGCOMM 2024 Conference*. 661–678.

[148] Han Zhang, Akram Bin Sediq, Ali Afana, and Melike Erol-Kantarci. 2024. Generative AI-in-the-loop: Integrating LLMs and GPTs into the Next Generation Networks. *arXiv preprint arXiv:2406.04276* (2024).

[149] Jie Zhang, Haoyu Bu, Hui Wen, Yongji Liu, Haiqiang Fei, Rongrong Xi, Lun Li, Yun Yang, Hongsong Zhu, and Dan Meng. 2025. When llms meet cybersecurity: A systematic literature review. *Cybersecurity* 8, 1 (2025), 1–41.

[150] Peiyuan Zhang, Guangtao Zeng, Tianduo Wang, and Wei Lu. 2024. Tinyllama: An open-source small language model. *arXiv preprint arXiv:2401.02385* (2024).

[151] Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, et al. 2023. A survey of large language models. *arXiv preprint arXiv:2303.18223* (2023).

[152] Hao Zhou, Chengming Hu, Dun Yuan, Ye Yuan, Di Wu, Xi Chen, Hina Tabassum, and Xue Liu. 2024. Large Language Models (LLMs) for Wireless Networks: An Overview from the Prompt Engineering Perspective. *arXiv preprint arXiv:2411.04136* (2024).

[153] Hao Zhou, Chengming Hu, Ye Yuan, Yufei Cui, Yili Jin, Can Chen, Haolun Wu, Dun Yuan, Li Jiang, Di Wu, et al. 2024. Large language model (llm) for telecommunications: A comprehensive survey on principles, key techniques, and opportunities. *arXiv preprint arXiv:2405.10825* (2024).

[154] Honghe Zhou, Xin Huang, and Lin Deng. 2024. Enhancing Network Traffic Classification with Large Language Models. In *2024 IEEE International Conference on Big Data (BigData)*. IEEE, 7282–7291.