

# Mapping the Landscape of Generative AI in Network Monitoring and Management

Giampaolo Bovenzi, Francesco Cerasuolo, Domenico Ciunzio, *Senior Member, IEEE*,

Davide Di Monda, Idio Guarino, Antonio Montieri, Valerio Persico, Antonio Pescapé, *Senior Member, IEEE*

**Abstract**—Generative Artificial Intelligence (GenAI) models such as LLMs, GPTs, and Diffusion Models have recently gained widespread attention from both the research and the industrial communities. This survey explores their application in network monitoring and management, focusing on prominent use cases, as well as challenges and opportunities. We discuss how *network traffic generation and classification*, *network intrusion detection*, *networked system log analysis*, and *network digital assistance* can benefit from the use of GenAI models. Additionally, we provide an overview of the available GenAI models, datasets for large-scale training phases, and platforms for the development of such models. Finally, we discuss research directions that potentially mitigate the roadblocks to the adoption of GenAI for network monitoring and management. Our investigation aims to map the current landscape and pave the way for future research in leveraging GenAI for network monitoring and management.

**Index Terms**—Generative AI, Networking, LLM, GPT, Diffusion Models, Traffic Classification, Intrusion Detection.

## I. INTRODUCTION

**B**ECAUSE of breakthroughs achieved in the last decade, *Generative Artificial Intelligence (GenAI)* stands as one of the most important stepping stones toward the intelligence era. At its core, GenAI excels in (i) distilling features of complex data distributions (uncovering intricate patterns) and (ii) utilizing these features to generate new, similar, yet distinct data. This contrasts with the usual *discriminative Artificial Intelligence (AI)* models that focus on analyzing, interpreting, and classifying data to solve specific inference tasks. This two-fold ability (i.e., complex analysis and generation) positions GenAI as a crucial technology in advancing both scientific research and industrial applications. Accordingly, GenAI supports tools designed to generate new content—text, images, videos, and more—based on patterns and information learned from large datasets.

At a higher abstraction level, such capabilities showcase GenAI as a powerful tool to solve intelligence-level tasks that are common to different domains: content generation, data augmentation, conversational agents and question-answering tools, human-machine interactions, and automation. Noteworthy examples of novel GenAI models are represented by Large

Language Models (LLMs), Diffusion Models, and State Space Models (SSMs). To specify, LLMs are language models built on the Transformer architecture, and they are referred to as “large” due to their vast number of parameters. Hereinafter, we use the terms “LLM” and “Transformer” synonymously to indicate the AI model [1]. Notable examples for these novel GenAI solutions are represented by GPT and LLaMA for LLM, DALL-E and Stable Diffusion for Diffusion Models, and Mamba for SSM. These models have demonstrated significant commercial value and technical potential. They show notable reasoning, generalization, and emergent abilities in different applications, like text-to-text, text-to-image, and text-to-code. As a consequence of such potential, the global GenAI market stood at just under 45 billion USD at the end of 2023 (doubling its value compared to 2022), and forecasts indicate an impressive growth of  $\approx 20$  billion USD per-year through 2030 [2].

The rapid development of GenAI has been fueled by **three main drivers**: (i) the availability of *large-scale data corpora*; (ii) methodological advancements in the AI field, i.e., the shift toward *deep and foundational generative models*; (iii) technological innovations supporting model building, i.e., *high-performance massive Graphics Processing Units (GPUs)*. Notably, despite these drivers, only a few global stakeholders (to date) are capable of training GenAI models from scratch. Hence, pre-trained large models are beginning to be shared by the open-source part of the GenAI community.<sup>1</sup>

On the other side, recent networking research has focused on using Deep Learning (DL) to develop efficient tools for *Network Monitoring and Management (NMM)* to meet modern Internet traffic needs. In this respect, GenAI can empower intent-based and autonomous networks by automating the translation of user objectives into actionable network policies [3]. This allows networks to self-configure, self-optimize, and self-heal, improving responsiveness and resilience. By leveraging GenAI’s predictive capabilities, networks can indeed anticipate traffic patterns and issues, ensuring seamless operation. This reduces manual management complexity, accelerates innovation, and enhances user experience in a dynamically changing digital landscape. However, the full utilization of GenAI for NMM requires shifting from common text, audio, and image generation to network-focused synthetic content—fulfilling the concept of “AI-generated everything” [4]. Despite the interest in integrating GenAI into networks and the Internet (trying to echo similar breakthroughs obtained in verticals such as

G. Bovenzi, F. Cerasuolo, D. Ciunzio, A. Montieri, V. Persico, and A. Pescapé are with the Department of Electrical Engineering and Information Technologies (DIETI) at the University of Naples Federico II, Italy (name.surname@unina.it).

D. Di Monda is both with DIETI and IMT School for Advanced Studies, Lucca, Italy (davide.dimonda@imtlucca.it).

I. Guarino is with the Department of Computer Science at the University of Verona, Italy (idio.guarino@univr.it).

<sup>1</sup>See e.g., <https://llama.meta.com/>.

computer vision or Natural Language Processing (NLP)) to date, general deployment issues [5] and unique networking challenges remain [6].

#### A. Contributions and Survey Organization

This article deepens the technical understanding of GenAI within the context of NMM. Accordingly, the **main contributions** provided by this manuscript can be summarized as follows:

- we discuss the **motivation behind our “GenAI landscape mapping”** effort in the field of NMM, highlighting the shared interest in GenAI from different stakeholders, as well as the gap in the (quickly-evolving) scientific literature we aim to fill with our work (Sec. II);
- we present a **categorization of novel GenAI methods**, offering the necessary background to help readers understand the distinctive aspects of NMM-specific research efforts and applications (Sec. III);
- we offer a **use-case-centric viewpoint**, discussing each **practical NMM use case and its interplay with GenAI** (Sec. IV), along with a **model-centric viewpoint** (Sec. V) to obtain a nuanced perspective. In addition, for the newly-branded GenAI solutions, we detail the proposed modifications to reference GenAI architectures and their code availability.
- we provide a comprehensive view of the **public datasets** leveraged for GenAI model lifecycle and the **available computing platforms** that can support and accelerate the design of novel GenAI-based NMM solutions (Sec. VI);
- finally, we briefly wrap-up the current GenAI **limitations** and identify potential methodological/technological **enablers** for deploying it safely and at scale in the NMM field (Sec. VII).

Figure 1 outlines the organization of the present survey, sketching the details of the sections constituting the manuscript.

## II. MOTIVATION OF GENAI IN NETWORK MONITORING AND MANAGEMENT: CONTEXT AND RELATED WORKS

In this section, we examine the increasing interest from both public and private stakeholders in using GenAI to support NMM processes (Sec. II-A). Next, we discuss related surveys that analyze the impact of GenAI methods in the networking domain II-B. Finally, we outline the positioning and scope of this survey (Sec. II-C).

#### A. GenAI in Networking: Context

The huge and general interest in GenAI solutions also maps to the networking domain, where recent initiatives reflect the endeavors of several private and public stakeholders. Table I provides an overview of this interest, reporting the efforts of different stakeholders in the context of GenAI for NMM.

For instance, the current interest in GenAI is witnessed by the recent establishment of IEEE ComSoc Emerging Technology Initiative on Large Generative AI Models in Telecom (GenAINet) [13]. ACM SIGCOMM has already featured several online talks in which experts have discussed the huge

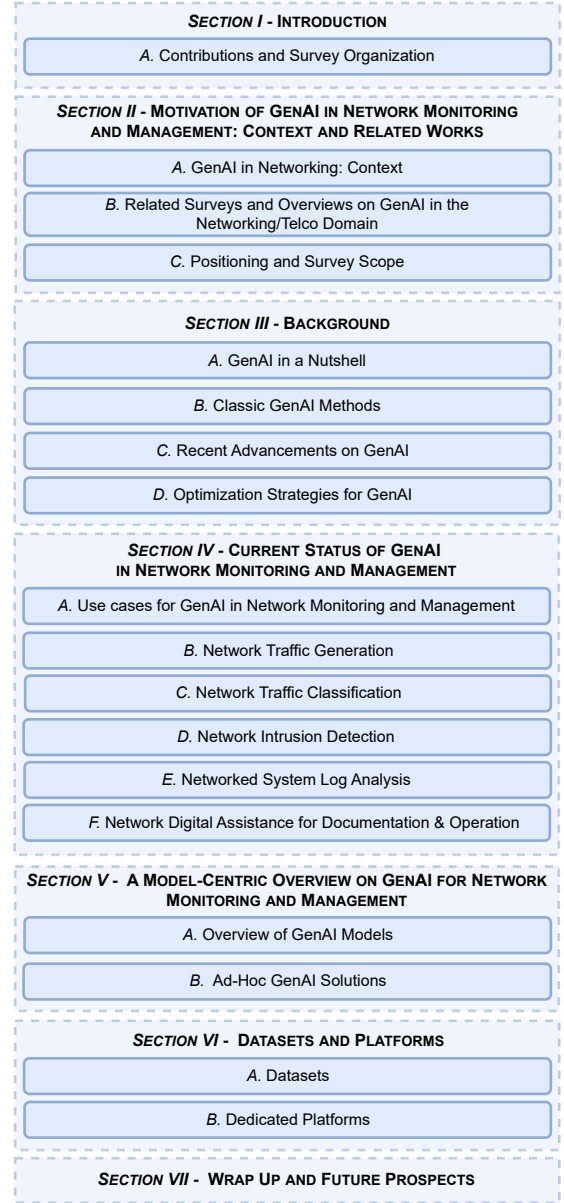














Figure 1. Survey organization.





interest in the application of GenAI to NMM (and, in general, to networking) [7]. Similarly, the International Telecommunication Union (ITU) via its initiative “AI for Good” is showcasing both industry- and academic-oriented viewpoints, as well as first LLM-based challenges [15]. Trending interest is also observed at the IETF, with a first side meeting entirely dedicated to the use of LLMs in networking [14]. It is further witnessed by the latest academic networking conferences and workshops that stably include the application of GenAI among the topics of their call, consistently seeking contributions in this direction (e.g., IEEE GLOBECOM 2024 will feature both dedicated workshops and symposia centered on GenAI).<sup>2</sup>

At the governmental level, the EU has launched a strategy for developing GenAI models over the past two years, highlighted by the EIC Accelerator funding program under

<sup>2</sup><https://globecom2024.ieee-globecom.org/call-papers>

Table I  
MAIN EFFORTS OF THE STAKEHOLDERS IN THE CONTEXT OF GENAI FOR NMM.

Stakeholder	Effort	Ref.
ACM SIGCOMM	 Scientific talks on LLMs for networking	[7]
AT&T	 Release of <i>Ask AT&amp;T</i> a GenAI digital assistant for employees and users	[8]
Cisco	 Release of a GenAI digital assistant as support to human decision-making	[9]
Ericsson	 Recognizing that GenAI will replace the traditional search process with a more intuitive and conversational experience	[10]
EU	 EIC Accelerator for developing GenAI, with a focus on transparency and smaller models	[11]
Huawei	 Release of <i>Net Master</i> , an LLM for network operations and maintenance tasks	[12]
IEEE ComSoc	 Platform for academia and industry for researching on GenAI for networking	[13]
IETF	 Scientific talks on LLMs for networking GenAI	[14]
ITU	 Showcase industry- and academic-viewpoints on LLMs	[15]
Nokia	 Recognizing the advantages of GenAI and identifying various networking use cases for its application	[16]
Telefonica	 Partnership with Microsoft to integrate GenAI in its networking ecosystem	[17]
TIM	 Integration of GenAI to support customer service and technical operations	[18]

: Academic Organization, : Organization of Multiple Entities, : Government Body/Agency, : Company.

the Horizon Europe framework aimed at supporting start-ups and small-medium enterprises [11]. Specifically, one of the 2024 challenges, “Human Centric Generative AI made in Europe” (50 million EUR budget), aims to promote a European human-centric approach to GenAI, addressing issues like transparency and trust, and seeking to (i) advance foundation language and multimodal frontier models, while also focusing on (ii) smaller foundation models with high performance in *specific domains*—like the case of this paper.

Network providers have also attempted to capitalize on the benefits of GenAI. For instance, Bell Labs acknowledges the benefits of GenAI, classifies several use cases (in areas such as customer care operations, network design, network performance and optimization, and testing), and envisions their expected role in shaping the future of organizations and functions of Telecom service providers [16]. Huawei has recently launched *Net Master* [12], an innovative network large model powered by GenAI that aims to enhance the efficiency of network operations and maintenance. This solution is trained using Huawei’s Pangu models (i.e., different foundation models tailored to different domains or specific use cases) and it is based on a 50-billion-level corpus and the experience of more than 10k networking experts. According to Ericsson, in the Telecom domain, the integration of GenAI capability to convert natural language to SQL and to execute complex SQL queries enables seamless interaction between users and data, replacing the traditional search process with a more intuitive and conversational experience [10]. This capability allows Telecom companies to empower users to effortlessly access and analyze data, easily supporting data-

driven decisions. Telefonica has partnered with Microsoft to integrate Azure GenAI in its digital ecosystem, enhancing its capabilities for key workflows, such as customer identity management or access to network Application Programming Interfaces (APIs) [17]. Similarly, AT&T has launched *Ask AT&T* a GenAI tool based on OpenAI’s ChatGPT, integrated within a secure AT&T-dedicated Azure environment [8]. This tool aims to enhance employees’ productivity by translating documents, optimizing network operations, updating legacy software, and improving customer support. On the same line, Cisco proposed its AI Assistant for accessing data at large scale to guide and inform human decision-making and enhance productivity while guaranteeing data protection and privacy [9]. Lastly, TIM is exploring the integration of GenAI across various sectors (e.g., marketing, customer care, network operations). The aim is to support customer service and technical operations through conversational interfaces, enhance document search and summarization, assist in code generation for IT tasks, and improve data analysis with natural language queries [18].

#### B. Related Surveys and Overviews on GenAI in the Networking/Telco Domain

Given the enormous hype surrounding GenAI techniques, a large number of recent surveys and tutorial-style studies aim to analyze and discuss their impact within the *wide domain of networking*. These works contribute to defining a rich but equally fragmented picture. Indeed, the available studies are characterized by different focuses, scopes, and depths in the provided pictures of the state of the art. Thus, they result in identifying (i) different vertical application fields, use cases, and networking tasks that can benefit from the (rapid) progress in GenAI, as well as (ii) different families of AI tools. Such studies and related aspects are summarized in Tab. II and briefly discussed in the following.

The majority of the works aim to analyze the role of GenAI in the fields of the **Internet of Things (IoT)** and/or **cybersecurity** [19–22, 29]. For instance, Sai et al. [19] explore the potential of combining GenAI with *IoT*, which enables the generation of synthetic data that can be used to train DL models to overcome data insufficiency or incompleteness in IoT systems. Ferrag et al. [29] provide a comprehensive survey of LLMs for *cybersecurity* identifying 9 application fields: threat detection and analysis, phishing detection and response, incident response, security automation, cyber forensics, chatbots, penetration testing, security protocol verification, and security training and awareness. Although the authors offer an in-depth analysis of the potentiality of LLMs for cybersecurity, the potential application fields they identify are not fully centered on networking and do not consider several promising LLM applications in this domain. Hassanin and Moustafa [20] overview the recent progress of LLMs in *cyber defense*, considering verticals that include threat intelligence, vulnerability assessment, network security, privacy preservation, and operations automation. Moreover, Halvorsen et al. [22] explore the application of GenAI for *intrusion detection* and discuss how GenAI can support penetration testing, supplementing datasets, or developing detection

Table II  
SURVEYS AND OVERVIEWS ON GENERATIVE ARTIFICIAL INTELLIGENCE IN RELATED FIELDS.

Work	Year	Focus	Vertical Application Fields, Use Cases, and Networking Tasks	AI Tools	#Surveyed Works
Sai et al. [19]	2024	IoT (data generation)	Synthetic Sensor Data Personalized Device Response Autonomous Control Cyber-threat Detection Predictive Maintenance Data Anonymization	GANs, VAEs, LLMs	15
Hassanin et al. [20]	2024	Cyber Defense	Threat Intelligence Vulnerability Assessment Network Security Privacy Preservation Operations Automation	LLMs	149
Alwahedi et al. [21]	2024	IoT Security	Cyber-threat Detection Lightweight Encryption Optimization Enhancing Access Control Identifying Vulnerability Automating and Enhancing Penetration Testing	ML†	62*
Halvorsen et al. [22]	2024	Intrusion Detection	Penetration Testing Supplementing Datasets Intrusion Detection Model Development	GANs, VAEs, LLMs	129
Zhou et al. [23]	2024	Telecommunications	Telecom-Domain Question Answering Troubleshooting Reports Generation Project Coding Network Configuration Network Attack Classification and Detection Telecom Text, Image, and Traffic Classification Performance Optimization Channel State Information Prediction Prediction-based Beamforming Traffic Load Prediction	LLMs	207
Celik et al. [24]	2024	Wireless telecommunication networks	Physical Layer Design Network Organization & Management Cross-layer Network Security Network Traffic Analytics Localization & Positioning	GANs, VAEs	303
Karapantelakis et al. [25]	2024	Mobile telecommunication networks	Improving aspects in RANs Mobile-network Management Requirements Engineering	GANs, VAEs, LLMs	117
Huang et al. [3]	2024	Networking	Threat Intelligence Vulnerability Assessment Network Security Privacy Preservation Operations Automation	LLMs	15
Liu et al. [26]	2024	Networking	Network Design Network Diagnosis Network Configuration Network Security	LLMs	15
Huang et al. [27]	2023	Networking	Network Design Network Diagnosis Network Configuration Network Security	LLMs	15
Chaccour et al. [28]	2024	Telecommunications	Network Operations (i.e., log analysis) Simplified Network Interfaces (i.e., APIs generation) Synthetic Data for Digital Twins DevOps and Software Lifecycle Management	GANs, LLMs	15
<i>This work</i>	2024	Network Monitoring and Management	Traffic Generation Traffic Classification Intrusion Detection Log Analysis Network Digital Assistance	LLMs, Diffusion Models, SSMs	189

\*: only a limited number of references is related to LLMs; †: LLMs only as a future trend.

models. They claim that both the training and test phases of intrusion systems benefit from GenAI. Alwahedi et al. [21] aim at providing a comprehensive overview of applying Machine Learning (ML) techniques for *IoT security*. In their future vision, the authors introduce the contribution of GenAI and LLMs to enhance IoT security—e.g., optimization of cyber threat detection, lightweight encryption, access control, vulnerability identification, and automated penetration testing. *Unfortunately, we underline that the GenAI applications and use cases discussed in the above surveys usually are not corroborated by existing state-of-the-art works.*

To the best of our knowledge, only a limited number of works [23, 24, 26, 27] aim at providing a **broader perspective** of GenAI in the **networking/telecommunication field**. In detail, Huang et al. [27] propose *ChatNet*, a domain-adapted network LLM framework with access to various external network tools. The authors discuss how LLMs promise to unify network intelligence through *natural language interfaces*. Specifically, they remark that domain adaptation of LLMs is paramount to fill the gap between natural language and network language and identify pre-training, fine-tuning, inference, and prompt engineering as the main enabling techniques. Liu et al. [26] provide a more condensed overview of the recent advances of LLMs in networking and present an abstract workflow to describe the fundamental process involved in applying LLM in such a domain, including task definition, data representation, prompt engineering, model evolution, tool integration, and validation. Interestingly, they remark that *network-specific LLMs* are expected to be more effective than using LLMs originally designed for general domains to perform network-related tasks. The works in [27] and [26] both identify network design, diagnosis, configuration, and security as the main vertical fields in networking impacted by LLMs. On the other hand, Zhou et al. [23] survey fundamentals, key techniques, and applications of LLM-enabled telecommunication networks. Specifically, they focus on four telecommunication scenarios: (i) generation problems, i.e., answering telecommunication-domain questions and generating troubleshooting reports, project coding, and network configuration; (ii) classification problems, i.e., network-attack, telecommunication-text, image, and traffic classification; (iii) network-performance optimization, i.e., automated reward function design to improve reinforcement learning applications; and (iv) prediction problems, i.e., prediction of channel state information and traffic load, and prediction-based beamforming. Karapantelakis et al. [25] focus on GenAI for *mobile telecommunication networks* and consider applications lying in verticals, such as optimizations in Radio Access Networks (RANs), network management, and requirements engineering. From the perspective of telco operations, Chaccour et al. [28] identify GenAI as a key to improving network operations such as predictive maintenance and real-time optimization. They discuss use cases of LLMs and GenAI for telco, including: (i) customer incident and trouble report management, proactive network management and repair, digital twin for network management, and intelligent network alert correlation—associated with LLMs; (ii) generating customized network configurations, creating dynamic service descriptions, and proactive

fault prediction and resolution—associated with GenAI solutions beyond LLMs, i.e., those performing content creation. Finally, Celik and Eltawil [24] focus on applying GenAI models within the domain of *wireless communications*. The authors provide a tutorial on GenAI models and a survey on their application across various wireless research areas, including: (i) physical layer design, (ii) network organization and management, (iii) network traffic analytics, (iv) cross-layer network security, and (v) localization and positioning. Specifically, in the domain of network traffic analytics, the authors focus on use cases such as network traffic generation, encrypted traffic classification, traffic prediction, and traffic morphing. In contrast, the exploration of GenAI networks security models is limited, with only a small portion addressing the enhancement of Network Intrusion Detection Systems (NIDSs), particularly in terms of improving their robustness. *As a final remark, we note that all the works surveyed in the areas of networking and telecommunications mainly utilize Generative Adversarial Networks (GANs), hence they only marginally cover the latest advancements involving more sophisticated techniques such as LLMs, Diffusion Models, and SSMs.*

For the sake of completeness, we mention that some works [30, 31] deepen *how the network is expected to support GenAI applications*, e.g., with focus on cloud-edge-mobile infrastructure and security & privacy concerns. We do not consider such research paths in our study but rather consider the opposite point of view, investigating *how GenAI can support network-related tasks*.

### C. Positioning and Survey Scope

In light of the rich but scattered literature scenario, we position the present work against the existing surveys and overviews in terms of the scope of the applications and tools considered, as well as the provided outcomes of the analyses.

To the best of our knowledge, none of the considered studies surveying the impact of recent advancements in GenAI primarily focuses on network monitoring and management. In fact, the studies that are primarily centered on networking [3, 26, 27] share a focus that is slightly close to ours. However, while envisioning the great potential of GenAI in networking, they lack a detailed survey and taxonomization of the current landscape, being aimed at *providing only a general overview based on the analysis of a very limited number of works* (indeed, these studies reference 15 papers each in their bibliography). On the other hand, the studies that provide a more systematic and in-depth analysis of the literature [23, 25] emphasize different facets of the communication networks, being oriented at capturing telecommunication aspects placed at lower layers in the communication stack—e.g., RAN improvement, mobile-network management, channel state information prediction, prediction-based beamforming. Hence, we believe *they provide a view that is complementary to ours*.

In this survey, we explore 5 use cases: (i) *network traffic generation*, (ii) *network traffic classification*, (iii) *network intrusion detection*, (iv) *networked system log analysis*, and (v) *network digital assistance*, which are crucial for network monitoring and management and are mostly overlooked in other such surveys.

Unlike all the related surveys, we perform an in-depth analysis of each mentioned use case aimed at identifying and providing taxonomies of the solutions proposed in the networking domain. Specifically, we report for each task the adopted GenAI architecture, its public availability, the input fed to the model, and the dataset leveraged for its pre-train or fine-tuning. Indeed, our study is intended for researchers and practitioners interested in capitalizing on the benefits of GenAI for network monitoring and management. Hence, we place a strong emphasis on the reproducibility of the proposals. Therefore, we also contribute to the taxonomization of the models used for each networking application we identify. While centered on the impact of the latest LLM wave, our study does not simply focus on LLM-based generative solutions—such as the majority of similar surveys [3, 20, 23, 26–28]. Instead, we analyze contributions that include the latest achievements based on Diffusion Models and SSMs, which are often overlooked in related surveys. On the other hand, we purposely exclude in our analysis generative algorithms such as GANs, Variational Autoencoders (VAEs), and normalizing-flows. These methods, while significant in past years, are considered less relevant compared to the latest advancements in GenAI.

### III. BACKGROUND

In this section, we first provide a formal description of GenAI (Sec. III-A). Then, we trace the evolution of GenAI models over time, from classic methods (Sec. III-B), proposed since 2013, to the most recent advancements of the present day (Sec. III-C). We end the section by describing the various strategies used to optimize GenAI models to deal with typical NMM use cases (Sec. III-D).

#### A. GenAI in a Nutshell

AI models can be classified into *discriminative* and *generative* models, according to the learning objective. The former makes predictions on unseen data by training on labeled data and thus can be used for various inference tasks. In contrast, generative models focus on synthesizing realistic content.

From a formal viewpoint, given a set of training samples  $\mathbf{x}_1, \dots, \mathbf{x}_N$  associated to an unknown data distribution  $p_d(\mathbf{x})$ , a GenAI technique learns a model to sample new (synthetic) data according to  $p_{mod}(\mathbf{x}) \approx p_d(\mathbf{x})$ . This can be accomplished by either two- or one-step approaches. In the former case, known as Explicit Density Estimation (EDE), the model first learns an explicit distribution  $p_{mod}(\mathbf{x}) \approx p_d(\mathbf{x})$  (in a tractable or approximate fashion), which is then used to sample new data. In the latter case, known as Implicit Density Estimation (IDE), the GenAI technique directly learns a model that can sample from  $p_{mod}(\mathbf{x}) \approx p_d(\mathbf{x})$  without explicitly defining it.

The design and use of GenAI have a long history in NMM: relevant methods include well-known Markov chains (tractable EDE) [32], but also VAEs (approximate EDE) [33], GANs (IDE) [34] and normalizing-flows (tractable EDE) [35]. Conversely, recent applications of generative models are *LLMs*—based on Transformer (and variants) or selective SSM—and Diffusion Models, which have represented a breakthrough

in the realism and complexity of the content generated. Transformer-based models enable parallelization and scalability, enhancing processing speed and contextual understanding through their self-attention mechanisms. Moreover, Diffusion Models offer several advantages over traditional generative models, such as VAEs and GANs, including better mode coverage and stability during training.

#### B. Classic GenAI Methods

Figure 2 reports the timeline of the development of GenAI, starting from VAEs (proposed in 2013 by Kingma and Welling [36] at the University of Amsterdam) until the latest models released by OpenAI in the second half of 2024, namely GPT-4o and its successive variants and evolutions (i.e., the lightweight GPT-4o mini, and the reasoning models o1-preview and o1-mini). We recall that *this survey focuses only on works that take advantage of the most recent advances in GenAI, specifically from the Transformer architecture onward*, which Google proposed in 2017 [37]. One motivation is that these solutions offer improved performance w.r.t. older solutions like VAEs, GANs, and normalizing-flows, e.g., Non-linear Independent Components Estimation (NICE) [38].

Moreover, this choice is justified by the impressive groundbreaking impact of these more sophisticated GenAI architectures across various fields, significantly improving generative tasks such as text generation, image synthesis, and multi-modal applications. However, we also report the *classic deep generative models*—viz., NICE, VAEs, and GANs with their variants and hybridizations—for context and completeness. Regarding the latter GenAI architectures—which fall outside our scope—we refer the reader to these prominent surveys for a deeper background and a detailed overview of their usage for networking-related use cases: [19, 22, 25, 39]. *The modern era of deep GenAI* started with VAEs at the end of 2013, GANs at mid-2014, and NICE at the end of 2014. These models were the first deep neural networks capable of learning generative models for complex data, such as images. In detail, VAEs introduced a structured and probabilistic approach to generative modeling with continuous latent spaces and improved training stability [36], while GANs presented a powerful adversarial framework that excels at generating high-quality and realistic data [40]. Then, NICE was the first model to implement normalizing flows using neural networks, leveraging them as invertible functions to transform data from a complex distribution to a simpler one [38].

Over time, diverse improvements to VAEs, GANs, and NICE have been proposed. Notably, the Conditional GAN (CGAN, 2014) enables controlled data generation by incorporating additional information into the generative process. This allows for a more targeted and context-specific output [41]. Additionally, the Deep Convolutional GAN (DCGAN, 2015) enhances the quality of generated images and improves the stability of the training process [42]. Moreover, hybrid architectures like VAE-GAN (2015) were proposed, integrating the structured latent space of VAEs into the adversarial training of GANs [43]. Lastly, the main evolution of NICE has been the Real Non-Volume Preserving (Real NVP) model—proposed in

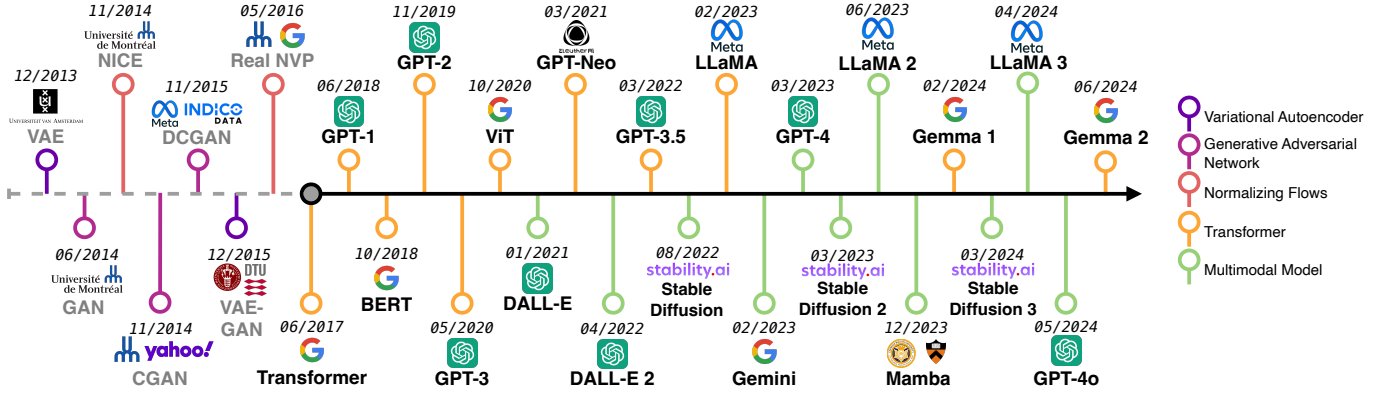


Figure 2. Timeline of GenAI development: while introducing VAN-based and GAN-based solutions, this work primarily focuses on developments from the Transformer onward.

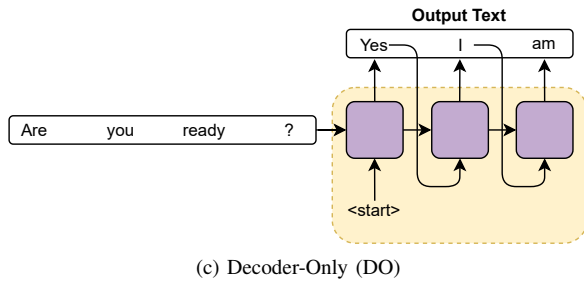
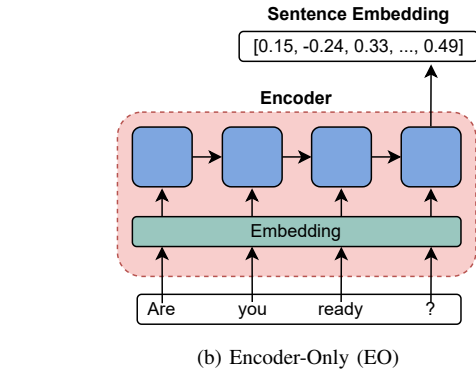
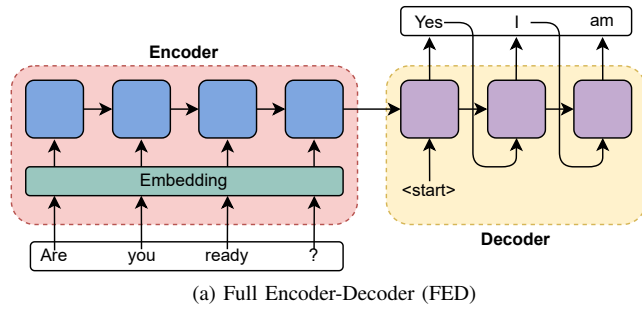


Figure 3. Overview of the general workflow of Transformer-based models: (a) Full Encoder-Decoder, (b) Encoder-Only, and (c) Decoder-Only.

mid-2016—that incorporates scale transformations, allowing the model to expand or contract regions of data rather than simply rotating or translating them, leading to more accurate and expressive generated contents [44].

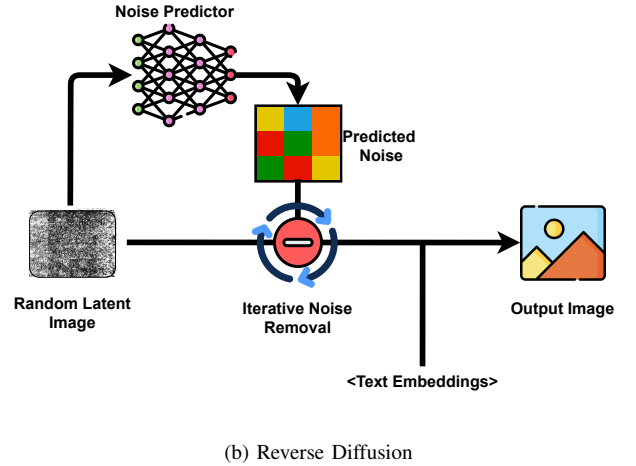
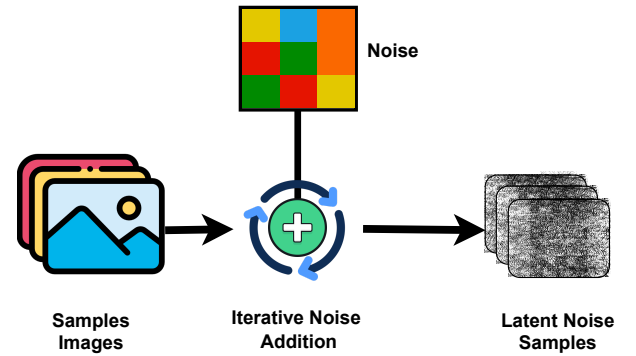


Figure 4. Overview of the general workflow of Diffusion Models, including (a) forward and (b) reverse diffusion processes.

### C. Recent Advancements on GenAI

Focusing on the most recent advancements in GenAI, namely from Transformer onward, we can identify five categories of architecture divided according to the nature of the underlying layers. We identify three variants of the Transformer architecture, namely the (i) *full encoder-decoder*, the (ii) *encoder-only*, and the (iii) *decoder-only* ar-

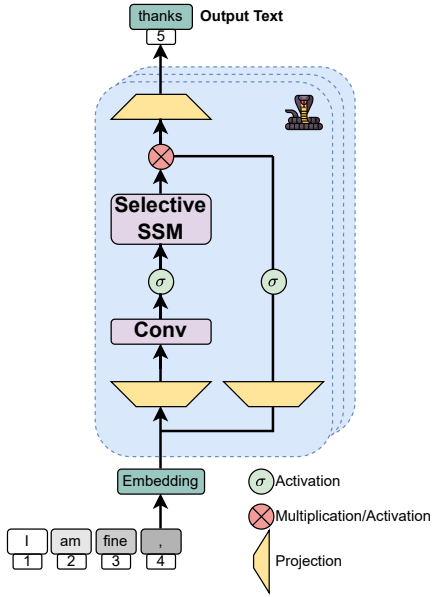


Figure 5. Overview of the general workflow of Mamba: the *Convolutional (Conv)* layer extracts relevant features from input data, focusing on spatial or temporal patterns; the *Selective SSM* layer filters and selects the most relevant latent states from the extracted features.

chitectures. Additionally, the other two categories are based on (iv) *diffusion processes* and (v) *state-space representations*, respectively. From a general perspective, the development of GenAI models has shifted in the last years toward a *foundational* nature definition [45]. Consequently, specific training strategies are commonly employed.

**Training Strategies for GenAI Models:** Three training strategies can be adopted for GenAI models. In the case of naive (a) *Monolithic Training*, the model is trained from scratch using a dataset tailored to the specific downstream task. More commonly, the training follows two sequential stages: (i) *Pre-Training*, where the GenAI model is pre-trained on a large corpus of data consisting of text, images, or other input modalities, in a self-supervised or semi-supervised manner. For instance, during this stage, a text-fed (resp. image-fed) model is instructed to predict masked words and the sequence of sentences (resp. to denoise or reconstruct the original picture). (ii) *Fine-Tuning*, where the GenAI model is then specialized for specific tasks (possibly by topping/modifying the architecture with task-specific layers). Specifically, the training parameters (or a portion of them) are jointly fine-tuned (exploiting the broader transfer learning concept), tailoring the model for the considered downstream task. Consequently, the training of a model can involve either (b) *Pre-Training & Fine-Tuning*, i.e., the model is first pre-trained on a large corpus of data (e.g., a networking corpus) and then fine-tuned with a dataset related to the downstream task, or (c) *Fine-Tuning Only*, i.e., an already pre-trained model is exclusively fine-tuned for the specific downstream task.

**Full Encoder-Decoder (FED):** This category includes the GenAI architectures that reflect the typical structure of the Transformer model [37]. The Transformer repre-

sents a revolutionary approach to solving sequence processing tasks. This architecture is entirely based on the self-attention mechanism rather than using recurrences—e.g., Recurrent Neural Network (RNN)—or convolutions—e.g., Convolutional Neural Network (CNN). This improvement enables the Transformer to efficiently parallelize computations and significantly reduce training times while achieving state-of-the-art results. In fact, traditional sequence processing models—such as RNNs and their variants like Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU)—perform computations around symbol positions in input and output sequences. This characteristic results in limited parallelization because it is inherently sequential, becoming a significant bottleneck for longer sequences. To cope with this drawback, the Transformer leverages self-attention mechanisms to model dependencies between different positions in a sequence, regardless of their distance.

In general, the Transformer architecture consists of an encoder-decoder structure. The general workflow of a FED is depicted in Figure 3a. The *encoder* comprises a stack of identical layers, each with two sub-layers, namely a multi-head self-attention and a position-wise fully connected feed-forward network. The *decoder* is similar to the encoder but includes an additional sub-layer that performs masked multi-head attention over the encoder’s output. It also modifies the self-attention sub-layer to prevent positions from attending to subsequent positions, ensuring the auto-regressive property. Both encoder and decoder are designed with residual connections for each sub-layer followed by layer normalization.

Going into detail, the Transformer uses multi-head attention to allow the model to learn information from different representation subspaces jointly. In fact, instead of having a single attention function, the model linearly projects queries, keys, and values multiple times with different learned projections and performs the attention function in parallel. This process enhances the model’s ability to focus on different parts of the input sequence. Moreover, since the Transformer lacks the inherent sequential order provided by recurrence, it introduces positional encodings to inject information on the position of tokens in the sequence. These encodings are added to the input embeddings at the bottom of the encoder and decoder stacks, enabling the model to understand the sequence order.

The Transformer is usually leveraged for sequence-to-sequence tasks, i.e., when the input and the output are both sequences. Examples of applications are translation (from one language to another), summarization (condensation of documents), and text generation (based on given prompts). Notable architectures that fall into this category are XLNet, T5, Gemini, Mistral, and Zephyr.

**Encoder-Only (EO):** Compared to FED, EO models only leverage the encoder unit of the FED architecture (as depicted in Figure 3b). EO is designed to model bidirectional relationships between tokens in an input sequence, generating either a vector representation for each token or a single vector summarizing the entire sentence. This architecture is well-suited for tasks focused on text understanding and analysis rather than generation.

Among the EO models, BERT [46]—Bidirectional Encoder Representations from Transformers—is the most representative and has served as the basis for many subsequent advances of this type of architecture. Developed by researchers at Google AI Language, BERT consists of multiple layers of bidirectional Transformer encoders. BERT has been designed with a key innovation: it pre-trains deep bidirectional representations from the unlabeled text by joint conditioning on both left and right contexts in all layers. This procedure allows BERT to capture richer linguistic information, and it contrasts with models like OpenAI GPTs and ELMo (an LSTM-based architecture), which are unidirectional and do not fully leverage the bidirectional context. In detail, the bidirectional training of BERT is achieved through a Masked Language Model (MLM) objective. The MLM randomly masks some tokens in the input sequence and predicts them using the context provided by the remaining tokens on both sides. This method allows BERT to capture the context from both directions. To further enhance its understanding of context and sentence relationships, BERT uses a next-sentence prediction task during pre-training. This involves predicting whether a given sentence B follows sentence A in the original text, allowing the model to learn how sentences relate to each other.

The ability of BERT to understand the context from both directions and the effectiveness of its pre-training tasks enables it to achieve superior performance across a wide range of NLP tasks. BERT has been designed for language understanding, i.e., encoding the input text in relation to its context for various subsequent tasks. Examples of applications are text classification (e.g., spam detection), question answering, and text similarity (e.g., semantic search). Other notable architectures in this category are variants of BERT, such as BERTiny, RoBERTa, DistilRoBERTa, and ViT (Vision Transformer).

**Decoder-Only (DO):** These models exploit only the decoder component of the FED architecture, as shown in Figure 3c. DO models are designed for autoregressive text generation, predicting the next token based on previous tokens, thereby producing the output one token at a time.

Among them, Generative Pre-trained Transformers (GPTs) [47] are the most prominent, spearheading advancements in the field of NLP starting with their first variant named GPT-1. Subsequent improvements of GPTs, including GPT-2, GPT-3, GPT-3.5, GPT-3.5 turbo, GPT-4, and GPT-4o, have dramatically increased the model size and the scale of pre-training data and have included multi-modality (text and images) from GPT-4 onward. GPT-4, with its estimated 1.7 trillion parameters<sup>3</sup>, exemplifies the trend towards larger models and has achieved state-of-the-art results across a wide range of benchmarks without task-specific fine-tuning.

GPTs are commonly leveraged for autoregressive text generation, i.e., generating text tokens conditioned on the previous token. Examples of applications are text generation, language modeling (e.g., autocompletion), and conversational AI or chatbots (e.g., ChatGPT and Copilot). Notable architectures in

this category are Falcon, LLaMA, Phi, Gemma and improvements of GPT-1, from GPT-2 to GPT-4o.

**Sequential Denoising Process (SDP):** Ho et al. [48] proposed a class of generative models, named Diffusion (probabilistic) Models, that describes the process by which particles, information, or other entities spread through a medium over time. In recent years, the Diffusion Model (Figure 4) has found successful applications in computer vision, as well as in audio, bioinformatics, and agent-based systems.

The core idea involves defining a *forward diffusion* process (Figure 4a) that gradually adds noise to the data, transforming them into a simpler distribution, typically Gaussian noise. The corresponding *reverse diffusion* process (Figure 4b) is then learned to map the noisy data back to the original data distribution. The elegance of Diffusion Models lies in their theoretical foundation, which leverages concepts from Markov chains (when diffusion is performed in discrete time) or stochastic differential equations (when diffusion is performed in continuous time). This foundation allows for a rigorous treatment of the model's behavior and facilitates efficient training and sampling algorithms (through sophisticated sampling acceleration techniques). The resulting models, such as the Denoising Diffusion Probabilistic Model (DDPM) and score-based generative models, have demonstrated remarkable capabilities in generating high-quality synthetic data. Because of the explicit definition of the forward/reverse diffusion process and the objective used to learn them (i.e., a generalized evidence lower-bound [49]), these models fall within the approximate EDE category. In summary, Diffusion Models have been used for high-quality data generation through iterative denoising. Examples of applications include visual and signal data processing tasks, such as image generation, audio synthesis, and video generation, contrasting with previous categories that primarily involve language and textual data-generation tasks. A notable model falling into this category is Stable Diffusion.

**Selective and Structured State Space Models (SSMs):** Proposed by Gu and Dao [50], Mamba represents a significant advancement in sequence modeling, introducing a new class of selective and structured SSMs designed to overcome the limitations of existing architectures like Transformers in handling very long input sequences. As depicted in Figure 5, Mamba integrates selective and structured SSMs into a streamlined neural network architecture that avoids traditional attention mechanisms, achieving *fast inference and linear scaling with sequence length*.

In fact, while Transformers have become the backbone of many foundation models due to their effective self-attention mechanism, they suffer from *quadratic scaling* w.r.t. sequence length, limiting their efficiency on long sequences. Subquadratic-time architectures, including linear attention and structured-only SSMs, have attempted to address these inefficiencies but have failed to perform in critical modalities such as language.

Mamba's first core innovation lies in making (a part of) the parameters of SSMs dependent on the input. This mechanism allows the model to *selectively propagate or forget information*

<sup>3</sup><https://the-decoder.com/gpt-4-has-a-trillion-parameters/>

based on the current token, i.e., the model can focus on relevant information or discard irrelevant or outdated information as needed. This selective mechanism enables Mamba to handle discrete modalities effectively, providing a significant advantage over previous SSMs. Secondly, Mamba reduces the number of trainable parameters by assuming a *structured* form for the SSM matrices defining the information propagation. Thirdly, to maintain efficiency, Mamba employs a (i) hardware-aware algorithm that computes the model recurrently without materializing the expanded state in the GPU memory (leveraging fast memory hierarchies) and (ii) a parallel scan algorithm to accelerate the recursive computation of relevant quantities. This approach ensures linear scaling in sequence length and high throughput on modern hardware.

In summary, Mamba integrates selective and structured SSMs into a simplified neural network architecture that omits attention and MultiLayer Perceptron (MLP) blocks. This streamlined design, inspired by previous SSM architectures, offers fast training and inference with high performance in various data modalities, including language, audio, and genomics. Mamba is designed to process and model sequences efficiently, making it ideal for applications that require handling long sequences and achieving high computational efficiency.

#### D. Optimization Strategies for GenAI

In this section, we discuss the optimization strategies that have been proposed for GenAI solutions, focusing on those that have been used in NMM use cases. Broadly, optimization strategies can be divided into two categories [51], involving methods for (i) *parameter-efficient fine-tuning* and (ii) *post-training quantization*.

**Parameter-Efficient Fine-Tuning (PEFT):** These methods enhance the adaptation of pre-trained GenAI models to the target downstream task. The primary objective is to minimize the computational resources required for fine-tuning while preserving inference performance.

Among recent advances proposed to optimize the fine-tuning step of GenAI solutions, the state-of-the-art approach named Low-Rank Adaptation (LoRA) [52] has been recently leveraged for traffic generation purposes [53]. LoRA capitalizes on the intuition that changes in model weights during adaptation have a low “intrinsic rank”. In other words, the idea is that adapting a model to a new task does not require very complex changes, which can be efficiently represented using fewer adjustments. In detail, LoRA optimizes the fine-tuning by using rank decomposition matrices, specifically targeting the change in dense layers during training while keeping the main pre-trained weights frozen. Thus, this method allows for efficient task adaptation by replacing some model components—i.e., parts of the weight matrices—with small low-rank matrices. This substitution reduces the need to recalculate gradients and memorize optimizer states. The way in which LoRA is designed ensures that there is no extra delay introduced during inference. Moreover, LoRA is compatible with other optimization techniques, such as post-training quantization [54].

**Post-Training Quantization (PTQ):** These methods aim to reduce the computational complexity and memory footprint of GenAI models by casting the model parameters into lower precision formats *after training*. Quantization facilitates faster inference and more efficient deployment on various hardware.

One of the most complete methods for enforcing PTQ that has been recently proposed is named GPT-Generated Unified Format (GGUF) [55]. GGUF is a generalized file format that has recently been adopted by [56] to enforce post-training quantization for network digital assistance purposes, namely for networking standards question answering. GGUF has been proposed to reduce—in large LLMs—the precision of the weights and activations of the model by converting real numbers to integers, e.g., 32-bit floating-point to 8-bit. GGUF has been devised with two key features in mind: *quantization-aware kernel optimization* and *extensibility*. On the one hand, GGUF does not simply apply quantization to the model weights but also provides kernel optimization functionalities that consider the quantization process. This characteristic is fundamental in avoiding an inference performance decrease due to blind quantization. On the other hand, GGUF has been designed to overcome the limits of its predecessor GGML<sup>4</sup>, which lacks mechanisms to incorporate additional model information or add new features. Therefore, GGUF allows the integration of new features into the file format while ensuring compatibility with models deployed in older GGUF formats, thus preserving backward compatibility for newer versions.

In general, GGUF provides several key functionalities, including single file deployment, improved model loading and saving speeds, and intuitive design and detailed information storage that facilitate extensibility. Together, these functionalities enable a more efficient and user-friendly experience in handling LLMs.

## IV. CURRENT STATUS OF GENAI IN NETWORK MONITORING AND MANAGEMENT

This section provides an overview of the current status of Generative Artificial Intelligence (GenAI) in the Network Monitoring and Management (NMM) context. Accordingly, Sec. IV-A outlines the five NMM use cases where GenAI is currently being utilized, along with a description of the benefits GenAI offers for each. Then, Secs. IV-B-IV-F dissect the works that employ GenAI for each use case.

### A. Use Cases for GenAI in Network Monitoring and Management

GenAI is actively used to address NMM use cases in various networking domains. We have identified five key use cases where it is currently employed: (i) *Network Traffic Generation*, (ii) *Network Traffic Classification*, (iii) *Network Intrusion Detection*, (iv) *Networked System Log Analysis*, and (v) *Network Digital Assistance for Documentation & Configuration*. Such use cases are summarized in Figure 6 along with the acronyms

<sup>4</sup>GGML is an ML library created by Georgi Gerganov, which is why it is named “GGML”. In addition to offering low-level ML primitives, such as tensor types, GGML also defines a binary format for distributing LLMs.

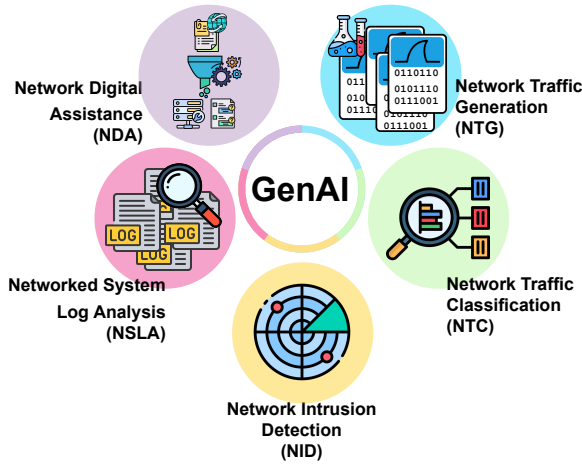


Figure 6. Overview of NMM use cases, leveraging GenAI, explored in this survey. These include Network Traffic Generation (NTG), Network Traffic Classification (NTC), Network Intrusion Detection (NID), Networked System Log Analysis (NSLA), and Network Digital Assistance (NDA).

we use in this work. Below, we provide detailed descriptions for each use case.

**Network Traffic Generation (NTG)** refers to the process of creating synthetic network data, ranging from the generation of (bi-)flow statistical features or sequence of features extracted from the packets within a (bi-)flow (e.g., packet size, inter-arrival time, and packet direction), to the generation of the entire PCAP trace.<sup>5</sup>

NTG is crucial for network traffic analysis from various applicative perspectives. It enables the simulation of different scenarios to (stress) test network infrastructure and services, validate security measures (e.g., for automated penetration tests), and augment training data for improving Machine Learning (ML) models performance and generalization capabilities. The key challenge of NTG is producing high-fidelity synthetic network samples that closely resemble real traffic. Hence, a critical aspect is the validation of the synthetic traffic generated, since both *effectiveness* (in enhancing ML or Deep Learning (DL) models performance) and *validity* (in simulating with high fidelity the real traffic) are desiderata of the synthetic network traffic generation task [39].

*How NTG can benefit from GenAI:* GenAI can significantly enhance the NTG task by leveraging its ability to understand and mimic natural language patterns, thus modeling network traffic as the “language of the Internet”. Accordingly, it helps in generating realistic protocol sequences and user interactions, producing high-quality synthetic traffic that mirrors the diverse and complex traffic patterns of real environments [53, 57].

**Network Traffic Classification (NTC)** aims to categorize network traffic represented by various *Traffic Objects (TOs)* such as packets, bursts, flows, biflows, or sessions. This process may include identifying the protocol (especially when nonstandard transport ports are used), the name of the ap-

plication (e.g., YouTube, Netflix, Facebook), or the type of service (e.g., streaming, web browsing, VoIP) that generated the traffic. Generally, NTC involves modeling target network traffic classes (e.g., by using ML or DL algorithms) and differentiating the traffic into one of these target classes.

Since NTC can identify user behaviors and predict traffic categories, it is crucial in enhancing network management operations. By applying rules based on NTC results, network management can be adapted to address the specific needs of the network, optimizing the handling of different types of traffic. The main challenges affecting NTC include the limited availability of high-quality data to train effective models and the poor generalization capabilities shown by the state-of-the-art NTC techniques [58–60].

*How NTC can benefit from GenAI:* GenAI can significantly enhance NTC through advanced contextual awareness and pattern recognition capabilities of pre-trained models. These models leverage large unlabeled datasets to learn unbiased data representations, which can be easily transferred to various downstream tasks by fine-tuning on limited labeled data. The killer idea can be the modeling of network traffic as a language, namely the language of machine-2-machine communication. Thus, GenAI can produce highly versatile pre-trained models that, due to their high generalizability, can be adapted to solve different NTC tasks with minimal effort, eliminating the need to train new models from scratch for each task [61, 62].

**Network Intrusion Detection (NID)** aims to identify anomalous or malicious traffic traversing the network. Specifically, NID focuses on monitoring the traffic exchanged between connected entities (e.g., mobile devices, computers, servers) to secure them. Its main objective is to detect anomalous behaviors that may be related to security threats or intrusions by analyzing the exchanged traffic.

NID is crucial in identifying malicious activities by distinguishing legitimate (viz., benign) traffic from potentially harmful (viz., malicious) traffic. Moreover, it can be used even to identify specific attack traffic [63]. These operations enable prompt response to threats and minimize potential damage to network infrastructure and its users.

NID should be seen as a specialization of NTC when dealing with supervised multiclass or binary classification (viz., misuse detection). In this context, NID leverages techniques common to NTC to identify types of attacks based on knowledge extracted from labeled training data. However, NID also encompasses Anomaly Detection (AD), which involves identifying outliers or abnormal behaviors that deviate from the norm. This process is typically addressed via out-of-distribution detection or one-class classification methodologies. For these reasons, we treated it as a separate use case in this survey also due to its importance, dedicated modeling solutions, and extensive related literature.

*How NID can benefit from GenAI:* Similarly to NTC, GenAI can enhance NID through contextual awareness, enabling the detection of anomalies by understanding the context of network events over time. Its adaptability by means of transfer learning allows rapid adaptation to new threats, while semantic

<sup>5</sup>In this survey, we do not cover sensor data generation, such as temperature or pressure measurements, since this task involves modeling physical phenomena rather than actual network traffic.

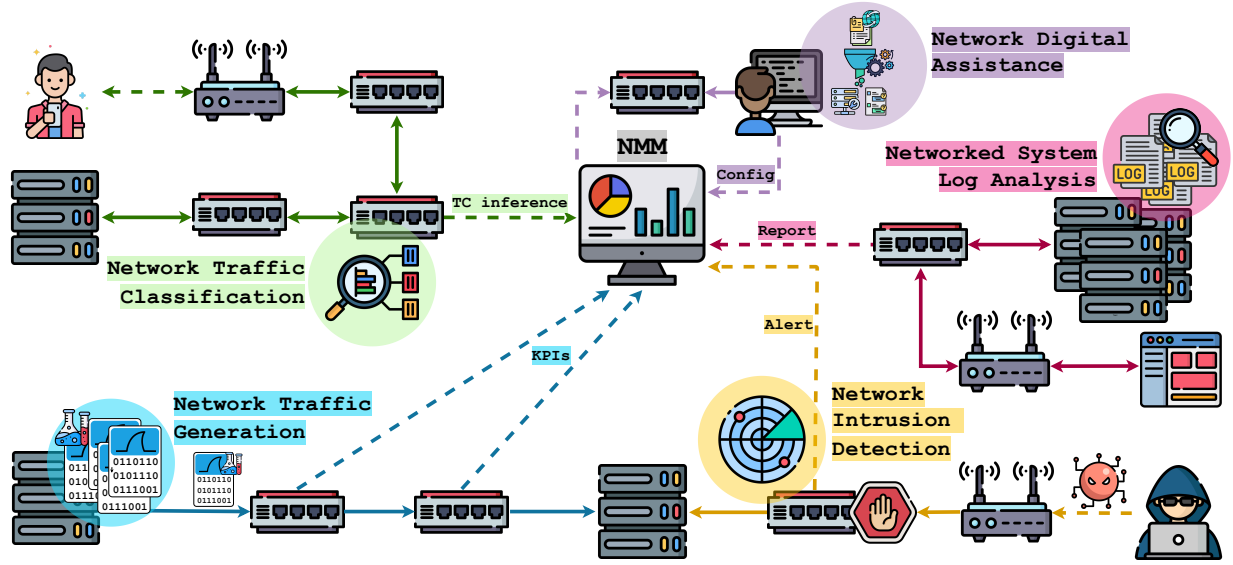


Figure 7. Possible interactions among NTC, NTG, NDA, NSLA, and NID to enhance network management efficiency, highlighting the key pathways and feedback mechanisms among the different entities.

analysis identifies unusual command sequences. In addition, complex pattern recognition and unsupervised learning help detect subtle deviations and unknown threats [64, 65].

**Networked System Log Analysis (NSLA)** refers to solutions that automate the extraction of knowledge from network or system logs to summarize them (i.e., to identify key elements) or to detect anomalies (viz., log anomaly detection). Network and system logs typically consist of semi-structured text/records of data that collect network or system events. Specifically, the logs considered in this work pertain to network-related applications, such as web or email servers, or related to network entities, such as network managers.

NSLA is crucial for enhancing security by identifying unauthorized access and potential security breaches. It ensures system reliability by providing hints to identify performance bottlenecks or diagnosing the root cause of the fault. Moreover, it improves software quality through log debugging or ensuring software robustness. It optimizes operations by analyzing user behaviors or auditing activities and helps maintain compliance across various domains (e.g., supporting predictive maintenance) [66].

*How NSLA can benefit from GenAI:* Leveraging modern and advanced Large Language Models (LLMs), GenAI can efficiently parse, interpret, and summarize log data written in natural language. It extracts significant events and patterns through semantic analysis, enhancing the understanding of log data [67, 68].

**Network Digital Assistance for Documentation & Operation (NDA)** (briefly, Network Digital Assistance) focuses on monitoring and controlling network operations to ensure efficient and reliable performance. Specifically, NDA aims to maintain network reliability and availability, optimize network performance, ensure security, and enable efficient resource scheduling. Moreover, it is fundamental for reducing

downtime, preventing data leaks, and ensuring uninterrupted service delivery. Hence, NDA is crucial for various applications. It facilitates interoperability issues in heterogeneous network environments characterized by multi-layer and multi-vendor infrastructures. NDA also supports advanced networking frameworks like Software Defined Networking (SDN) and simplifies the management of ever-growing Internet of Things (IoT) environments. In this context, resource provisioning, device configuration, network monitoring, and software update management are essential for reducing energy consumption and strengthening the security of IoT devices that are resource-constrained and insecure-by-design [69–71].

*How NDA can benefit from GenAI:* GenAI allows extensive automation for network operations. Specifically, LLMs provide a natural language interface that simplifies the retrieval of complex information in networking standards and documents crucial for NDA. This interface also facilitates the management of various network software and hardware, enhancing operational efficiency [72].

**Relations among NMM use cases:** In general, these five use cases are strongly related. Together, they improve the monitoring and control of network operations, leading to improved network performance, reliability, and security. From a broad perspective, NMM acts as the system's actuator, leveraging the outputs of other tasks to make informed decisions and optimize network performance. A graphical representation of this interaction is shown in Figure 7, highlighting the links between the various use cases.

Specifically, NTG creates synthetic traffic that can be used to evaluate potential network configurations before deployment—i.e., NDA. By using reliable and accurate traffic generators, which can produce traffic that closely resembles real traffic, NTG can provide valuable insights into the behavior of network equipment in pseudo-real operating environ-

ments. Additionally, by generating various types of synthetic traffic, both benign and malicious, NTG can be utilized to assess, in different contexts, the response of classification and intrusion detection systems—i.e., NTC and NID, respectively. Moreover, it can enhance their performance and generalizability by augmenting training data for ML and DL models.

Conversely, NTC and NID offer insights into the (real) traffic traversing the network, enabling performance improvements (through specific traffic prioritization or routing rules) and enhancing security (by applying filtering rules to block anomalous or malicious traffic). Thus, NTC and NID can facilitate online (re)configurations through NDA. Additionally, data-driven NTC and NID systems trained on real traffic data can provide valuable feedback on the quality of the synthetic traffic generated—i.e., validating NTG.

Finally, while NSLA shares similarities with NTC and NID mechanisms, it focuses on analyzing log data related to network equipment (e.g., servers and routers) and the services provided (e.g., web pages) rather than network traffic. Therefore, NSLA supports network management operations by providing summaries and insights from logs and by identifying anomalies in the operation of network equipment and services. This information can then be used to adjust the behavior of these network equipment and services (i.e., NDA).

**From network traffic to GenAI:** Figure 8 details the generic pipeline for the use cases. Specifically, NTG, NTC, and NID leverage the same kind of input (i.e., network traffic) to pre-train and fine-tune the GenAI model. Conversely, NSLA and NDA typically employ pre-trained models, which are then fine-tuned with specific log data or different network documents to adapt to the task at hand.

GenAI approaches are not designed to ingest network traffic directly. Instead, approaches based on LLMs or Diffusion Models are typically designed to process data in a text-based or image-like format. Therefore, for use cases involving direct processing of network traffic (i.e., NTG, NTC, and NID), traffic data need to be transformed into a text-based or image-like representation before using it as input to the GenAI model. This transformation is performed via *Datagram-to-Token* and *Datagram-to-Image* operations, respectively.

**Datagram-to-Token:** Approaches based on LLMs typically employ a *Datagram-to-Token* method to convert encrypted traffic into pattern-preserved token units for pre-training [61]. This method involves segmenting traffic into packets and representing their characteristics as word-like tokens, similar to natural language processing. When packets are grouped into traffic objects, such as (bi)flows or bursts, special tokens are required to mark the packet boundaries within these traffic objects (e.g., common values for these special tokens are [SEP], [MSK], [PAD], and [PKT]). Additionally, the type of features extracted from the traffic object (or from each packet belonging to it) may require further preprocessing before being converted to tokens (e.g., conversion into raw bytes, anonymization, or quantization).

**Datagram-to-Image:** Approaches based on Diffusion Models typically employ a *Datagram-to-Image* method to convert

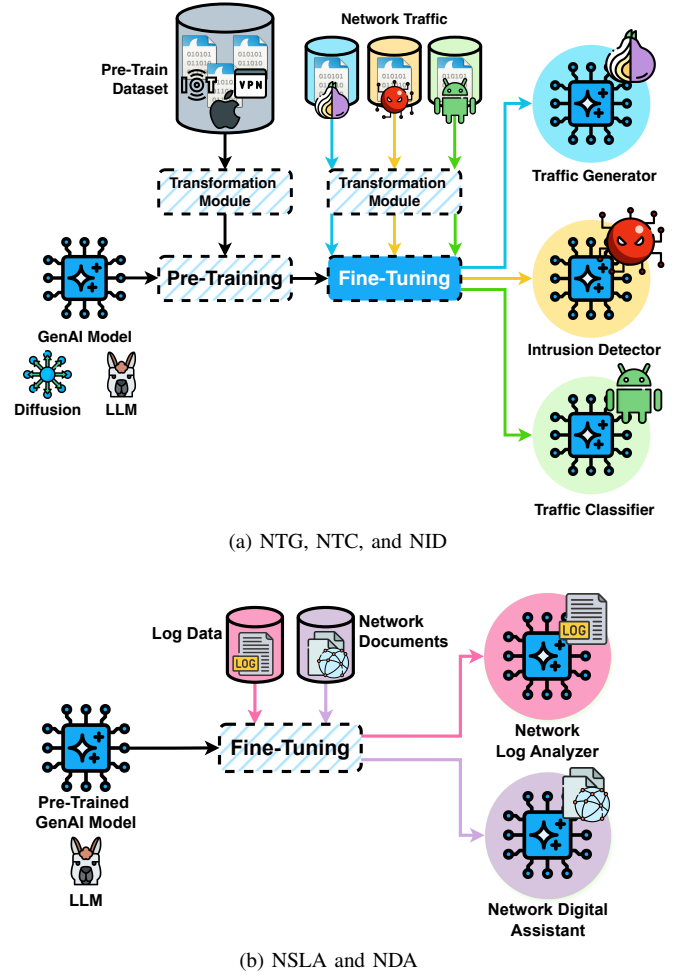


Figure 8. Pipeline for NMM use cases with GenAI models, detailing the process for (a) NTG, NTC, NID, and (b) NSLA and NDA. Dashed-line blocks indicate optional stages.

encrypted traffic into image representations [53]. Two main Datagram-to-Image variants are commonly exploited based on the features to be used: (i) raw-bytes-to-image and (ii) features-to-image. The former involves translating network traffic into standardized bits, where each bit corresponds to a packet header field bit. The encoded sequence of packets is then formed into a matrix, which is interpreted as an image (e.g., *nPrint* format [73]). Conversely, when the model input is a time series of packet features (e.g., packet sizes or inter-packet times), different transformations can be applied to encode the time series as an image, such as *FlowPic* [74] or *Gramian Angular Summation Field (GASF)* [75].

Hereinafter, we provide a detailed overview of the existing literature for each NMM use case.

### B. Network Traffic Generation

**Definition:** NTG entails creating synthetic data that accurately replicate real-world network traffic patterns and behaviors.

Table III summarizes works addressing NTG with the GenAI model, published since 2021. Notably, some studies are not explicitly focused on generating synthetic network traffic. In fact, they also tackle tasks related to *traffic understanding*,

Table III  
WORKS DEALING WITH NTG THROUGH GENAI MODELS (IN CHRONOLOGICAL ORDER).

Paper	Year	GenAI Model		GenAI Train		Traffic Input			Networking Dataset			Generated Data	Evaluation Metrics
		Name	Architecture	Technique	NetPT	TO	Data	Format	Monolithic Train or Pre-Train	Fine-Tuning	Fidelity Evaluation		
Bikmukhamedov and Nadeev [76]	2021	–	Lightweight GPT-2	MT	–	B	Packet Header Fields	Text	UNSW-IoT-Analytics Private† UNSWB-NB15 ISCXVPN2016	–	UNSW-IoT-Analytics Private†	Sequence of Packet Header Fields	KS
Meng et al. [57]	2023	NetGPT	GPT-2	PT&FT	⊗	F/P	Raw Packet Bytes	Text	ISCXVPN2016 USTC-TFC2016 CIRA-CIC-DoHBrw2020 PrivII 2021†	ISCXVPN2016 USTC-TFC2016 CIRA-CIC-DoHBrw2020 Cybermining-2023†	ISCXVPN2016 USTC-TFC2016 CIRA-CIC-DoHBrw2020 Cybermining-2023†	Packet Header Fields	JSD
Sivaroopan et al. [75]	2023	NetDiffus	DDPM	MT	–	T	Aggregated Traffic or Packet Header Fields	Image	Video Streaming† Deep Fingerprinting IoT Smart-Home†	–	Video Streaming† Deep Fingerprinting IoT Smart-Home†	Sequence of Aggregated Traffic or Packet Header Fields	FID, CA
Kholgh et al. [77]	2023	PAC-GPT	GPT-3	PT&FT	⊗	F	Packet Summaries	Text	–	TON_IoT	TON_IoT	Python code for flow traffic generation	SR
Jiang et al. [53]	2024	NetDiffusion	Stable Diffusion 1.5	PT&FT	⊗	B	Raw Packet Bytes	image	–	Private†	Private†	Sequence of Raw Packet Bytes	JSD, TVD, HD, CA
Wang et al. [78]	2024	LENS	T5 1.1	PT&FT	⊗	F	Raw Packet Bytes	Text	ISCXVPN2016 ISCXTor2016 USTC-TFC2016 CIRA-CIC-DoHBrw2020 CIC-IoT-2023	ISCXVPN2016 ISCXTor2016 USTC-TFC2016 CIRA-CIC-DoHBrw2020 CIC-IoT-2023 Cross-Platform	ISCXVPN2016 ISCXTor2016 USTC-TFC2016 CIRA-CIC-DoHBrw2020 CIC-IoT-2023 Cross-Platform	Packet Header Fields	JSD, TVD
Qu et al. [79]	2024	TrafficGPT	GPT-based	MT	–	F/P	Raw Packet Bytes	Text	ISCXVPN2016 USTC-TFC2016 Cross-Platform ISCXTor2016 CIRA-CIC-DoHBrw2020 CIC-IoT-2022	–	HTTP flow† DNS flow† TLS flow†	Packet Header Fields	JSD
Chu et al. [80]	2024	–	Mamba	MT	–	F	Raw Packet Bytes	Text	Video Streaming† Video Conferencing† Social Media†	–	Video Streaming† Video Conferencing† Social Media†	Sequence of Raw Packet Bytes	JSD, TVD, HD
Zhang et al. [81]	2024	NetDiff	DDPM	MT	–	B	Per-Flow Counters/Stats	Numeric	LabeledFlows_2017 LabeledFlows_2019	–	LabeledFlows_2017 LabeledFlows_2019	Per-Flow Counters/Stats	JSD, TVD, CRPS CA, $R^2$
Li et al. [82]	2024	LW-Diff	Lightweight Diffusion Model	MT	–	F*	Raw Packet Bytes*	Text	USTC-TFC2016	–	USTC-TFC2016	Raw Packet Bytes*	CA, PR, REC, F1
Wolf et al. [83]	2024	–	GPT-2	MT	–	F	Aggregated Traffic or Packet Header Fields	Text	CSE-CIC-IDS2018 TON_IoT UNSWB-NB15	–	CSE-CIC-IDS2018 TON_IoT UNSWB-NB15	Aggregated Traffic or Packet Header Fields	JSD, MAE-Corr* FPR, F1

*GenAI Model Architecture*: DDPM - Denoising Diffusion Probabilistic Model; *GenAI Train*: MT - Monolithic Train, PT&FT - Pre-Train & Fine-Tuning, NetPT - Networking Pre-Train, ⊗ - Present, ⊖ - Absent; *Traffic Input*: TO - Traffic Object; T - Trace, B - Bidirectional Flow, F - Flow, P - Packet;

*Networking Dataset*: †: Private dataset, [76] considers two train scenarios: w/ all the datasets or w/o those in *italic*;

*Evaluation Metrics*: HD - Hellinger Distance, JSD - Jensen-Shannon Divergence, KS - Kolmogorov-Smirnov, TVD - Total Variation Distance, FID - Frechet Inception Distance, CRPS - Continuous Ranked Probability Score, MAE-Corr - Mean Absolute Error of Correlation Matrices, CA - Classification Accuracy, PR - Precision, REC - Recall, F1 - F1-score,  $R^2$  - Coefficient of Determination, SR - Success Rate; \*: based on the attribute type, the correlation among attributes is computed via: (i) the Pearson Correlation Coefficient, (ii) the Uncertainty Coefficient, or (iii) the Correlation Ratio;

\*: Information marked with \* is not explicitly reported in the reference work but has been inferred.

such as NTC and NID [57, 78, 79] by opportunistically fine-tuning GenAI models.

Most of the reviewed works employ LLMs (e.g., GPTs and T5) [57, 76–79] while other leverage Diffusion Models (e.g., Stable Diffusion) [53, 75, 81]. The sole exception is the work in [80] that uses Mamba. The column “**GenAI Model - Architecture**” outlines the architecture used by each study.

As for the TO, the considered works segment the traffic into packets [57, 79], unidirectional [57, 77–80, 83] or bidirectional [53, 76, 81] flows, or even consider the whole network traffic trace [75]. As input data for GenAI models (column “**Traffic Input - Data**”), almost all works employ raw packet bytes [53, 57, 78–80], optionally performing IP masking operations [53, 78, 79]. Other studies leverage the sequence of packet header fields such as Packet Sizes (PSs) and Inter Arrival Times (IATs) [76], or incorporate Packet Directions (DIRs), IATs, and optionally PSs [75]. Conversely, Sivaroopan et al. [75] and Wolf et al. [83] also employ aggregated metrics (e.g., forward/backward volume and packet count, and flow duration), while Kholgh and Kostakos [77] exploit packet summary extracted with the Linux’s `tcpdump` tool. Finally, Zhang et al. [81] leverage the sequences of traffic statistics related to, for instance, the forward/backward number of packets, IATs, and traffic volumes.

As described in Sec. IV-A, LLMs and Diffusion Models are not inherently designed for handling network traffic. Therefore, to employ these architectures, the input data need to be formatted either in a text-based [57, 76–81, 83] or in an image-like [53, 75] representation (see column “**Traffic Input - Format**”). For datagram-to-text conversion, these

approaches use complex *tokenization mechanisms*, to preserve the complex hierarchical structure (i.e., spatial and temporal properties of packets within a flow) of real network traffic. Additionally, Qu et al. [79] encodes time information into tokens, enabling the model to generate timestamp intervals for a comprehensive representation of PCAP file data. On the other hand, Diffusion Models need an *encoding-decoding strategy* to transform network traffic data into an image format and subsequently convert them back to the original traffic format. To perform this operation, Sivaroopan et al. [75] exploit the Gramian Angular Summation Field (GASF) [84] method while Jiang et al. [53] use nPrint [73].

As described in Sec. III-C, the training procedure of GenAI models relies on two different strategies (see column “**GenAI Train**”): (i) the naive *Monolithic Training (MT)* or (ii) *Pre-Training & Fine-Tuning (PT&FT)*. Most of the studies in Tab. III adopt the MT approach [75, 76, 79–83], training the models with heterogeneous traffic data from networking datasets including ISCXVPN2016, USTC-TFC2016, and CIRA-CIC-DoHBrw2020. For models exploiting PT&FT, while the fine-tuning is always performed on networking datasets, the column “**NetPT**” indicates whether the pre-training phase includes a networking-specific corpus. In fact, Kholgh and Kostakos [77] and Jiang et al. [53] adapt pre-trained models not originally trained on networking data to networking-specific tasks. Specifically, Kholgh and Kostakos [77] leverage OpenAI’s GPT-3, pre-trained on a mix of publicly available and licensed Internet text, while Jiang et al. [53] use Stability AI’s Stable Diffusion 1.5, trained on the LAION dataset. Differently, Meng et al. [57] and Wang et al. [78]

perform the pre-training phase on networking datasets and successive fine-tuning on specific downstream datasets. For more details on such datasets, please refer to Tab. X.

NTG approaches can also be categorized based on their output (column “**Generated Object**”). Some studies focus on generating specific packet fields, such as IP addresses, ports, and packet sizes [57, 78, 79, 83] or aggregated traffic features, such as traffic volume, flow duration, and number of forward/backward packets [81, 83]. It should be noted that these methods can also be applied iteratively to generate entire flows, akin to constructing sentences from individual words in the Natural Language Processing (NLP) domain. Differently, other approaches natively generate sequences of packet header fields (e.g., PS, IAT, and DIR) [75, 76], aggregated traffic features [75], or raw traffic bytes of entire biflows [53, 80]. Interestingly, Kholgh and Kostakos [77] provide Python code for interacting with the Scapy library<sup>6</sup> to generate traffic that matches the input desiderata. This solution is more akin to traffic replay than generation.

Finally, to evaluate the fidelity and realism of generated data, the considered works leverage two different kinds of metrics (column “**Evaluation Metrics**”): (i) divergence/fidelity metrics, (ii) ML-based classification accuracy of related downstream tasks, or (iii) success rate. In the former case, the most common metrics are the Jensen-Shannon Divergence (JSD) [53, 57, 78–81, 83] and the Total Variation Distance (TVD) [53, 78, 80, 81] that quantify the similarity and the maximum difference between two distributions, respectively. Both metrics range from 0 (identical distributions) to 1 (completely different distributions) and are commonly evaluated together. Hence, lower values signify synthetic traffic more similar to the real one. Conversely, in the case of ML-based evaluation, the synthetic traffic is used during the training or evaluation phases of different ML models targeted for various downstream tasks. In the case of a traffic classification task, the variation in accuracy [53, 75, 80–82], F1-score [82, 83], Precision, or Recall [82] is taken as a measure of the quality of the generated data. Zhang et al. [81] also evaluate the generated data through traffic prediction assessed via  $R^2$ . In addition, Wolf et al. [83] analyze the False Positive Rate (FPR) of a model (viz. discriminator) specifically trained to distinguish between real and synthetic traffic, with higher FPR values indicating the difficulty in distinguishing between the two types of traffic, thereby reflecting their similarity. Lastly, Kholgh and Kostakos [77] evaluate the synthetic traffic generated by their Python code based on the Success Rate (SR), quantifying the proportion of successfully sent packets out of the total generated ones.

### C. Network Traffic Classification

**Definition:** NTC involves categorizing network traffic based on various attributes such as protocols, services, and application types.

Table IV summarizes the approaches employing LLMs to address NTC. It details the model type, traffic input, classification tasks, and training and evaluation datasets.

All the reviewed works leverage LLMs, such as Transformer, BERT, GPT-2, and Mamba, adapted from other domains (e.g., NLP) to the traffic context (column “**Architecture**”) and optionally modify some architectural elements. For instance, [85] propose a hierarchical Transformer model to process data at multiple levels of granularity (e.g., intra- and inter-bursts). The adaptation involves transforming network traffic data into a text-based representation, followed by tokenization, allowing models to learn the complex characteristics of network traffic directly. Notably, for the tokenization process, considered works leverage *Datagram2Token* [61, 62, 85, 87], *SentencePiece* [57, 78, 79], *WordPiece* [78], or *Byte-Pair Encoding* [86].

As shown, most of the approaches consider the flow as the TO, while Sarabi et al. [86] leverage the network service in terms of destination IP and port (column “**TO**”). The input data typically includes the header and payload of the network layer [57, 62, 78, 79, 88], or only the network-layer [61] and transport-layer payload [87]. Conversely, Guthula et al. [85] leverage fields from both the transport and application layers along with metadata at both packet and burst levels, while [86] focus on HTTP messages (column “**Traffic Input Data**”).

Furthermore, all the reviewed works include a pre-training stage for the LLM architecture (column “**NetPT**”), through a self-supervised learning approach typically involving two tasks: (a) *Masked Burst Model* to capture the relationships between different datagram bytes within the same burst, and (b) *Same-origin Burst Prediction* to model the transmission relationships between preceding and subsequent bursts. Then, the resulting model is fine-tuned by adapting the pre-trained model to various traffic classification tasks and adjusting its parameters to optimize performance on the labeled data. These processes are performed through the use of different datasets. Primarily, most of the works leverage ISCXPV-2016 [57, 61, 62, 78, 79, 87], USTCTFC-2016 [57, 62, 79], CIC-DoHBrw-2020 [57, 78, 79], or ISCXTor-2016 [62, 78, 79]. Additionally, also traffic from Network Intrusion Detection System (NIDS) (e.g., CIC-IDS-2017, CIC-IDS-2018) and IoT devices (e.g., CIC-IoT-2022, CIC-IoT-2023) is included to pre-train LLMs [61, 62, 78, 79, 87] (column “**Pre-Training/Fine-Tuning Datasets**”).

The reviewed works address different NTC tasks, differing in granularity and classification types. Specifically, most approaches focus on classifying the service [61, 62, 78, 85, 87, 88] (ref. **Serv.**) or application [57, 61, 62, 78, 79, 87, 88] (ref. **App.**) generating the traffic. Other works focus on detecting VPN-/Tor-encapsulated traffic [57, 78, 88] (ref. **Encaps.**). To this end, these approaches primarily use datasets incorporating diversified traffic on multiple levels (e.g., ISCXPV-2016 and ISCXTor-2016) during both the pre-training and fine-tuning stages. Conversely, only a few works focus on classifying DNS queries using the DoH protocol [78] (ref. **Query Met.**). Noteworthy, some studies introduce datasets specifically for evaluation purposes, which are not used during the pre-training and fine-tuning phases. Examples include Cross Platform [61, 78, 79, 88] and CSTNET-TLS1.3 [61].

<sup>6</sup><https://scapy.net/>

Table IV  
WORKS DEALING WITH NTC THROUGH GENAI MODELS (IN CHRONOLOGICAL ORDER).

Paper	Year	GenAI Model			Pre-Training Dataset	Traffic Input		Downstream Tasks					
		Name	Architecture	NetPT		TO	Data	Dataset	Encaps.	Serv.	App.	Query Met.	Device
Lin et al. [61]	2022	ET-BERT	BERT	⊙	ISCVN2016 CIC-IDS2017 CSTNET†	F	L3-PAY	ISCVN2016 ISCVN2016 CSTNET-TLS1.3 Cross-Platform	○ ○ ○ ○	● ○ ○ ○	● ● ● ●	○ ○ ○ ○	○ ○ ○ ○
Meng et al. [57]	2023	NetGPT	GPT-2	⊙	ISCVN2016 USTC-TFC2016 CIRA-CIC-DoHBrw2020 PrivII 2021†	F/P	L3-H+PAY	ISCVN2016 CIRA-CIC-DoHBrw2020	● ○	○ ○	● ○	○ ●	○ ○
Guthula et al. [85]	2023	netFound	Transformer	⊙	Private†	F	L4/L5 Fields + Metadata	Private†	○	●	○	○	○
Sarabi et al. [86]	2023	—	RoBERTa	⊙	CensysBQ	S	HTTP Messages	CensysBQ	○	○	○	○	●
Wang et al. [78]	2024	LENS	T5 1.1	⊙	ISCVN2016 ISCVN2016 USTC-TFC2016 CIRA-CIC-DoHBrw2020 CIC-IdT-2023	F	L3-H+PAY	ISCVN2016 ISCVN2016 CIRA-CIC-DoHBrw2020 Cross-Platform	● ● ○ ○	● ● ○ ○	● ○ ● ○	○ ○ ○ ●	○ ○ ○ ○
Qu et al. [79]	2024	TrafficGPT	GPT-based	⊙	ISCVN2016 USTC-TFC2016 Cross-Platform ISCVN2016 CIRA-CIC-DoHBrw2020 CIC-IdT-2022	F/P	L3-H+PAY	ISCVN2016 USTC-TFC2016* Cross-Platform	○ ○ ○ ○	○ ○ ○ ○	● ● ● ●	○ ○ ○ ○	○ ○ ○ ○
Wang et al. [62]	2024	NetMamba	Mamba	⊙	ISCVN2016 ISCVN2016 USTC-TFC2016 Cross-Platform CIC-IdT-2022	F	L3-H+PAY	ISCVN2016 ISCVN2016 Cross-Platform	○ ○ ○	● ● ○	○ ○ ●	○ ○ ○	○ ○ ○
Liu et al. [87]	2024	LAMBERT	BERT	⊙	ISCVN2016 CIC-IDS2017 CSE-CIC-IDS2018 CSTNET-TLS1.3	F	L4-PAY	ISCVN2016 CSTNET-TLS1.3	○ ○	● ○	● ●	○ ○	○ ○
Li et al. [88]	2024	—	ALBERT	⊙	CSTNET-TLS1.3	F	L3-H+PAY	ISCVN2016 CSTNET-TLS1.3 EduTLS† Cross-Platform	● ○ ○ ○	○ ○ ● ○	○ ● ○ ●	○ ○ ○ ○	○ ○ ○ ○

GenAI Model Architecture: **NetPT** - Networking Pre-Train, ⊙ - Present, ⊗ - Absent; **Traffic-Input**: **TO** - Traffic Object: **F** - Flow, **P** - Packet, **S** - Service, **L3/L4** - ISO/OSI Network/Transport Layer, **H/PAY** - Header/Payload Bytes, **Fields** - Header Fields; † - Private Data; **Downstream Tasks**: **Encaps.** - Traffic Encapsulation Identification, **Serv.** - Traffic Service Classification, **App.** - Application Classification, **Query Met.** - Query Metrics Classification, **Device** - Device Classification;  
**Note**: \* TrafficGPT by Qu et al. [79] has been naively applied to a security dataset (i.e., USTC-TFC2016) without mentioning the NID use case. Accordingly, we treat this work as NTG and NTC, without including it in the NID use case.

## D. Network Intrusion Detection

**Definition:** NID is an umbrella term that covers network security tasks that aim to identify and recognize malicious behavior from network traffic and collectively contribute to the design of so-called NIDS.

Accordingly, in this section, we review works that propose GenAI solutions ending up under the NID use-case umbrella. Such works are described in Tab. V alongside four main views, namely (i) “**Downstream Task**” details, (ii) the leveraged “**GenAI Model**” characteristics, (iii) the “**Traffic Input**” fed, and (iv) “**Pre-Training**” and “**Fine-Tuning Datasets**”.

First, we categorize the literature based on the “**Downstream Task**”. Indeed NIDS can be taxonomized based on their modeling objective, namely Misuse Detection (MD) or Network Anomaly Detection (NAD). MD relies on recognizing both normal (benign) and anomalous (malicious) behaviors in a supervised fashion, whereas NAD focuses only on normal network traffic and identifies anomalies as deviations from legitimate behavior. Within MD, we also distinguish between Binary Misuse Detection (bMD) and Multi-class Misuse Detection (mMD). The main difference is that mMD can identify specific attack types (and thus enable attack-tailored countermeasures), whereas bMD simply distinguishes between legitimate and malicious traffic. By looking at the reviewed literature, only the work by Nam et al. [89] performs NAD, while the remaining studies are divided between bMD [57, 64, 78, 91, 92, 95, 96, 98, 99] and mMD [61, 62, 65, 78, 85, 90, 94, 97, 100, 101], with [93]

performing both.

Secondly, regarding the “**GenAI Model**”, the most common choice falls in basic Transformers or ViT (i.e., full encoder-decoder category) [61, 85, 91, 93, 94, 98, 99], then BERT or variants follow (i.e., encoder-only category) [64, 65, 90, 92, 95, 96, 100], and minor attention is posed on GPT-based solutions (i.e., decoder-only category) [57, 89, 97, 101]. Other models, such as Google’s T5 [78] and Mamba [62] are also explored. In detail, Transformers and ViT are commonly used “as-is” in many studies [91, 94, 98]. Wang et al. [99] enhance the transformer encoder’s training phase by incorporating a contrastive loss term, and the encoder output is subsequently fed into a transformer decoder for sample reconstruction. Similarly, BERT is often employed with minimal modifications [65, 90, 92, 96]. Ghourabi [95] extends BERT by proposing a framework that utilizes it alongside *LightGBM* (Light Gradient Boosting Machine, an open-source distributed gradient boosting framework). Manocchio et al. [64] introduce FlowTransformer, which integrates BERT and GPT-3. Their architecture includes a shallow encoder and decoder, GPT as a deep decoder, and BERT as a deep encoder, with a MultiLayer Perceptron (MLP) for classification. Regarding studies based on GPT, Nam et al. [89] combine two GPT-1 networks in a bi-directional manner, employing a forward and a backward GPT followed by a dense layer and softmax. Ali and Kostakos [97] use GPT-3.5 turbo for explainability purposes alongside various eXplainable Artificial Intelligence (XAI) techniques. Lastly, Melcias et al. [101] leverage GPT-1 for data augmen-

Table V  
WORKS DEALING WITH NID THROUGH GENAI MODELS (IN CHRONOLOGICAL ORDER).

Paper	Year	GenAI Model			Traffic Input		Datasets		Downstream Task
		Name	Architecture	NetPT	TO	Data	Pre-Training	Fine-Tuning	
Nam et al. [89]	2021	–	GPT-1	⊗	B	CAN ID sequences	–	Hyundai Avante CN7	NAD
Yu et al. [90]	2021	–	BERT	⊗	AS	APT characteristics	–	Power Grid Data	mMD
Li et al. [91]	2022	ESeT	Transformer	⊗	F	L4-PAY, Packet-level byte encoded features, Flow-level frequency domain features	–	CIC-IDS2017 CSE-CIC-IDS2018	bMD
Seyyar et al. [92]	2022	–	BERT	⊗	B	HTTP requests	–	CSIC 2010 FWAF HttpParams	bMD
Ho et al. [93]	2022	–	ViT	⊗	B	Stats	–	CIC-IDS2017 UNSW-NB15	bMD, mMD
Wu et al. [94]	2022	RTIDS	Transformer	⊗	B	Stats	–	CIC-IDS2017 CIC-DDoS2019	mMD
Ghourabi [95]	2022	–	BERT	⊗	B	Stats	–	ECU-IoTH TON_IoT Edge-IIoTset	bMD
Lin et al. [61]	2022	ET-BERT	Transformer	⊙	F	L3-PAY	ISCXVPN2016 CIC-IDS2017 CSTNET†	USTC-TFC2016	mMD
Lai [96]	2023	–	BERT	⊗	F	Stats	–	ISCX NSL-KDD	bMD
Ali et al. [97]	2023	HuntGPT	GPT-3.5 turbo	⊗	F	Stats	–	KDD'99	mMD
Ullah et al. [98]	2023	TNN-IDS	Transformer	⊙	P, F, B	Stats	MQTT-IoT-IDS2020	MQTT-IoT-IDS2020	bMD
Wang et al. [99]	2023	RUIDS	Transformer	⊗	BoF	Stats	–	KDD'99 UNSW-NB15 CIC-IDS2017 (Friday) CIC-IDS2017 (Wednesday)	bMD
Meng et al. [57]	2023	NetGPT	GPT-2	⊙	F	L3-H+PAY	ISCXVPN2016 USTC-TFC2016 CIRA-CIC-DoHBrw2020 PrivII 2021†	USTC-TFC2016 CIRA-CIC-DoHBrw2020 Cybermining-2023†	bMD
Guthula et al. [85]	2023	netFound	Transformer	⊙	F	L4/L5 Fields + Metadata	Private†	CIC-IDS2017	mMD
Manocchio et al. [64]	2024	FlowTransformer	GPT-3 BERT	⊗	F	Stats	–	ISCX NSL-KDD UNSW-NB15 CIC-IDS2017 CSE-CIC-IDS2018 MQTT-IoT-IDS2020 TON_IoT	bMD
Wang et al. [78]	2024	LENS	T5 1.1	⊙	F	L3-H+PAY	ISCXVPN2016 ISCXTor-2016 USTC-TFC2016 CIRA-CIC-DoHBrw2020 CIC-IoT-2023	USTC-TFC2016 CIRA-CIC-DoHBrw2020 CIC-IoT-2023	bMD, mMD
Wang et al. [62]	2024	NetMamba	Mamba	⊙	F	L4-PAY	ISCXVPN2016 ISCXTor-2016 USTC-TFC2016 Cross-Platform CIC-IoT-2022	USTC-TFC2016 CIC-IoT-2022	mMD
Wang et al. [100]	2024	BT-TPF	BERT-of-Theseus	⊙	N.D.	Stats	CIC-IDS2017	CIC-IDS2017 TON_IoT	mMD
Melicias et al. [101]	2024	–	GPT-1	⊗	B	Stats	–	Edge-IIoTset	mMD
Ferrag et al. [65]	2024	SecurityBERT	BERT	⊗	F	L3/L4/L5 Fields	–	Edge-IIoTset	mMD

GenAI Model Architecture: **NetPT** - Networking Pre-Train, ⊙ - Present, ⊗ - Absent; **Traffic Input**: **TO** - Traffic Object; **AS** - Attack Sequence, **B** - Bidirectional Flow, **BoF** - Bag of Flows, **F** - Flow, **P** - Packet, **L3/L4/L5** - ISO/OSI Network/Transport Layer, **H/PAY** - Header/Payload Bytes, **Fields** - Header Fields; † - Private Data; **Downstream Tasks**: **NAD** - Network Anomaly Detection, **mMD** - multiclass Misuse Detection, **bMD** - binary Misuse Detection.

tation. They demonstrate that GPT-based methods can generate invalid data, leading to performance degradation in mMD.

Concerning the particular data format fed to models (see “**Traffic Input**” column), the majority of studies [64, 93–101] use pre-processed features, such as flow-based statistics typically derived from pre-processed datasets. Four works [89, 90, 92] utilize inputs specific to their particular domain, such as HTTP requests (for anomalous HTTP requests detection), Controller Area Network (CAN) ID sequences (for CAN intrusion detection) or Advanced Persistent Threat (APT) attack sequence (for IoT APT attack detection). The remaining studies use fields extracted from packet headers [65, 85], payload bytes [61, 62, 91], or both [57, 78].

One third of the works [57, 61, 62, 78, 85, 98, 100] pre-trains the models. The “**Pre-Training Dataset**” may belong to various domains such as: VPN traffic (i.e., ISCXVPN-2016 [57, 61, 62, 78]), Tor traffic (i.e., ISCXTor-2016 [62, 78]), and malicious traffic in IoT and non-IoT contexts (i.e., CIC-IDS2017 [61, 100], USTC-TFC2016 [57, 62, 78], CIRA-CIC-DoHBrw-2020 [78], and CIC-IoT-2022 [62, 78]).

Differently from the pre-training phase, for the downstream task, the datasets (see column “**Fine-Tuning Dataset**”) are exclusively related to the cybersecurity domain. The most common dataset is CIC-IDS2017 [64, 85, 91, 93, 94, 99, 100]. Other frequently used datasets include those for IoT-domain attacks (e.g., Edge-IIoTset [65, 95, 101], MQTT-IoT-IDS2020 [64, 98], TON\_IoT [64, 95, 100], and CIC-IoT-2022 [62]). Some studies rely on outdated datasets (i.e., KDD'99 [97, 99] and its improved version NSL-KDD [64, 96]) collected over 20 years ago, which no longer accurately reflect contemporary network traffic.

#### E. Networked System Log Analysis

**Definition:** NSLA tasks involve parsing logs to extract meaningful insights and information critical for maintaining and optimizing network operations.

Table VI provides a comprehensive overview of various research proposals employing GenAI for log analysis within the networking domain during 2021–24.

Table VI  
WORKS DEALING WITH NSLA THROUGH GENAI MODELS (IN CHRONOLOGICAL ORDER).

Paper	Year	Input Type	GenAI Model					Evaluation/Fine-Tuning Datasets	Downstream Task
			Name	Architecture	Fine-Tuned	Released	Goal		
Setianto et al. [102]	2021	Honeypot logs	GPT-2C	GPT-2	⊗	⊗	Q&A Parsing	CyberLab Honeynet	Parsing Unix commands in real-time
Ott et al. [103]	2021	Cloud environment logs	–	BERT, GPT-2, XLNet	⊗	⊗	Log Vectorization	Loghub (OpenStack)	LAD from preprocessed log performed via Bi-LSTM
Pan et al. [104]	2023	Supercomputer logs	RAGLog	GPT-3.5	⊗	⊗	LAD	Loghub (BGL, Thunderbird)	Detecting anomalies by leveraging a Retrieval Augmented Generation
Qi et al. [105]	2023	Supercomputer logs	LogGPT	GPT-1	⊗	⊗	Q&A LAD	Loghub (BGL), Spirit	Prompt-based LAD
Jiang et al. [106]	2023	Various log types	LILAC	GPT-3.5 turbo	⊗	⊗	Log Parsing	Loghub	Parse logs using in-context learning and adaptive cache
Ji et al. [107]	2023	Deployment framework and hardware network logs	–	GPT-2	⊗	⊗	LAD	Ada and Bob	Detect anomalies using semantic and sequential features with alarm strategy
Mudgal et al. [108]	2023	Cloud environment logs, supercomputer logs and prompts	–	GPT-3.5 turbo	⊗	⊗	Q&A Parsing, Analytics, Summarization	Loghub	Log parsing and summarization, API and LAD with ChatGPT
Sun et al. [109]	2023	Spring Boot apps, Kubernetes clusters, HTTP services logs	–	Generic LLMs	⊗	⊗	LAD	Spring Boot, Kubernetes, and HTTP services logs	Log management in cloud-native environments with real-time monitoring and LAD.
Han et al. [67]	2023	Supercomputers and distributed file system logs	LogGPT	GPT-2	⊗	⊗	LAD	Loghub (HDFS, BGL, Thunderbird)	LAD with GPT model fine-tuned using reinforcement learning
Vörös et al. [110]	2023	URLs	–	BERT, BERTiny, T5, GPT-3	⊗	⊗	URL Classification	Private security vendor dataset†	Web content filtering for maintaining network security and regulatory compliance
Boffa et al. [68]	2024	Unix shell and honeypot logs	LogPrécis	BERT, CodeBERT, CodeBERTa, GPT-3	⊗	⊗	Parsing and Analysis	NLP2Bash, HaaS†, CyberLab Honeynet, Polito†	Create an attack fingerprint assigning attacker tactics to each portion of a session
Balasubramanian et al. [111]	2024	Web server logs and prompt	CYAGENT	GPT-3.5 turbo, CodeT5	⊗	⊗	Log Summarization	Web Server Access Logs, Self-generated (Access Logs)†	Analyze and summarize logs, detect events and deliver cybersecurity information
Karlsen et al. [112]	2024	Supercomputer and web server logs	–	BERT, RoBERTa, DistilRoBERTa, GPT-2, GPT-Neo	⊗	⊗	LAD	Apache Web Server, CSIC 2010, ECML/PKDD 2007, Spirit Loghub (Thunderbird, BGL)	Unsupervised LAD with embedding compression via autoencoders and self-organizing map
Meyuhas et al. [113]	2024	Network traffic logs	–	RoBERTa, GPT-1	⊗	⊗	Device Function Labeling	IoT Sentinel, Censys, MUDIS, UNSW-IoT-Analytics, IoTFinder	AI-automated IoT labeling with vendor and device function
Tian et al. [114]	2024	DNS logs	Dom-BERT	BERT	⊗	⊗	DNS Logs Reconstruction	self-built†	Malicious DNS entry detection on logs
Almodovar et al. [115]	2024	Supercomputer and distributed file system logs	LogFIT	RoBERTa Longformer	⊗	⊗	LAD	Loghub (HDFS, BGL, Thunderbird)	Self-supervised LAD with fine-tuned LLM

Fine-Tuned/Released GenAI Model Architecture: ⊗ - Present, ⊖ - Absent; Evaluation/Fine-Tuning Datasets: † - Private Data.

As the input of GenAI models (column “**Input Type**”), considered works employ logs related to events and activities of operating system and applications in honeypots [102], cloud environments [103, 108], DNS [114], web servers [109, 111, 112], Kubernetes clusters [109], supercomputers [104, 108, 115], and Unix shell [68]. Differently, Meyuhas et al. [113] leverage network traffic log, which captures data from the flow of packets exchanged between IoT devices, while [110] employ URLs collected on firewalls and endpoints. Additional details on the dataset used for testing and optionally fine-tuning the GenAI model are provided in the “**Evaluation/Fine-Tuning Dataset**” column, with the majority of reviewed works leveraging Loghub datasets [67, 103–106, 108, 112, 115]. For details about datasets, please refer to Tab. X.

Furthermore, some works take as input also prompts to interact with GenAI models [102, 105, 108, 111]. Listing 1, reported in [108], provides an example of a possible prompt interaction for error and root cause identification within system logs.

```
Summarize the errors and warnings from these log messages
and identify the root cause.
[Sun Dec 04 04:52:49 2005] [notice] workerEnv.init() ok
/etc/httpd/conf/workers2.properties
[Sun Dec 04 04:52:49 2005] [notice] workerEnv.init() ok
/etc/httpd/conf/workers2.properties
[Sun Dec 04 04:52:52 2005] [error] mod_jk child workerEnv
in error state 7
```

Listing 1. Example of prompt.

All the considered works employ transformer-based models from different categories (see “**GenAI Architecture**” column and cf. Sec. III). As a *decoder-only* architecture, various GPT-based models are employed, such as GPT-2 [67, 102, 103, 107, 112], GPT-3 [68], GPT-3.5 [104] or GPT-3.5 turbo [106,

108, 111], GPT-Neo [112]. In contrast, as *encoder-only* architectures, BERT [68, 103, 110, 113, 114] is used, as well as its derived versions, such as RoBERTa [112, 113, 115], DistilRoBERTa [112], BERTiny [110], CodeBERT [68], and CodeBERTa [68]. Additionally, Balasubramanian et al. [111], Ott et al. [103], and [110] leverage *full encoder-decoder* models, i.e., T5, CodeT5, and XLNet which also include an auto-regressive module. Furthermore, Almodovar et al. [115] employ Longformer, which overcomes BERT limitations in handling sequences exceeding 512 tokens.

Most studies leverage publicly-available GenAI models [103–109, 113], while others fine-tune them [67, 68, 102, 110–112, 114, 115] (see “**Fine-Tuned**” column). Only some studies release the updated version of the model (see the “**Released**” column). Notably, Han et al. [67] leverage reinforcement learning strategy to fine-tune and adapt the GenAI model for Log-based Anomaly Detection (LAD), that is, identifying unusual (viz., anomalous) patterns or behavior in system logs that deviate from the normal ones.

The “**Downstream Task**” column details the specific task each study addresses, while the “**GenAI Goal**” column describes the functions that GenAI models perform to achieve these tasks.

On the one hand, Setianto et al. [102] and Jiang et al. [106] perform accurate real-time log-parsing. Although both use GenAI models to perform log parsing—which involves extracting structured information from unstructured log data—the former [102] employs a Q&A interaction with GenAI model, while the latter [106] leverages in-context learning and adaptive parsing cache. In-context learning optimizes the creation of diverse prompts, while adaptive parsing cache stores and updates parsed log templates to avoid redundant

queries and ensure accuracy.

On the other hand, a significant number of works focus on LAD [67, 68, 103–105, 107–109, 111, 112, 115]. Interestingly, four of these do not use GenAI models directly for LAD but as preprocessing components. Specifically, Mudgal and Wouhaybi [108] and Boffa et al. [68] perform log parsing for LAD. The former work uses prompt-based interactions with ChatGPT using pre-defined prompts, while the latter work creates an attack fingerprint and assigns attacker tactics within the *MITRE ATT&CK tactics*<sup>7</sup> to each session portion to reveal the attacker's goals. Ott et al. [103] focus on vectorization, converting log data into numerical vectors and performing LAD using nearest template matching to manage incomplete prior knowledge of log templates. Balasubramanian et al. [111] perform summarization, condensing log files into concise, human-readable formats for LAD. Among the works explicitly using GenAI models for LAD, Qi et al. [105], similarly to [108], use prompt-based interactions with ChatGPT, leveraging prompt-construction strategies. Similarly, Pan et al. [104] employ a Q&A strategy with log entries and best-matched retrieved entries from a database to determine whether a queried log entry is normal or not. Furthermore, Ji et al. [107] encode normal patterns and define an alarm strategy to filter out false positives based on statistical log data characteristics. Then, Almodovar et al. [115] leverage a self-supervised training strategy on normal log data to learn its linguistic and sequential patterns, thereby distinguishing it from malicious logs.

Lastly, Meyuhas et al. [113] use GenAI models to analyze network traffic and automatically classify IoT devices by vendor and function, offering insights into the traffic they generate. Tian and Li [114] examine DNS logs to discover malicious DNS entry—using GenAI models for log reconstruction—while in [110] web content filtering based on URLs is accomplished with GenAI models directly tackling URL classification.

#### F. Network Digital Assistance for Documentation & Operation

**Definition:** NDA refers to the process of administering, controlling, and optimizing network operations to ensure efficient functionality, performance, and security. NDA is pivotal due to the heterogeneous nature of networks, which often consist of diverse hardware and software components from different vendors. This diversity introduces complexity, making it challenging to navigate through various network configurations, protocols, and standards effectively [72].

Table VII provides a summary of the papers dealing with NDA, emphasizing the related key aspects. The publication dates of the works highlight the recent interest of the scientific community in this topic (2023 – 24). This interest has primarily converged on two downstream tasks, as seen from the “**Downstream Task Description**” column, which can be identified as follows:

- Using GenAI as virtual assistants to query standard documents in the networking/telecommunication domain,

thereby providing support to users (viz., *Network Digital Assistance for Documentation*).

- Employing GenAI as assistants for the network operative phase, e.g., handling network topologies, setup of device configurations, and infrastructure management (viz., *Network Digital Assistance for Operation*).

Detailing, 8 works [56, 116, 120–124, 126] focus on the *design of a virtual assistant specific to the telecommunication domain*. The authors of [120, 124] highlight the difficulty in analyzing and extracting information from standard documents in the telecommunication domain, as it involves identifying sources from multiple documents and related references. Thus, the authors investigate whether LLMs can be used as digital assistants for Q&A on standard documents. Ahmed et al. [123] extend the functionalities proposed in [120] to enrich the digital assistant's functionalities. In particular, they also include tasks related to text classification and summarization, as well as Q&A. Unlike the previous ones, in [56, 126], a digital assistant specifically designed for Third Generation Partnership Projects (3GPPs) standards is introduced. Unlike previous research focused on Q&A for standardized documents, the assistant proposed in [116] specifically targets user-support tasks. These include finding information about products and services, initial assistance with installation and configuration, and operational tasks like troubleshooting and performance monitoring. Duclos et al. [122] develop an LLM-based assistant designed to translate Request for Comments (RFCs) into a format compatible with Cryptographic Protocol Shapes Analyzer (CPSA).

The works in [117–119, 122, 125, 127] address the *design of a network operation assistant*. Contrary to the previous studies, these papers exhibit greater heterogeneity due to the inherent task diversity associated with the operational phase. Mani et al. [118] tackle the complexities of network topology and communication graph analysis by exploiting LLMs. They demonstrate how these models can be used to generate task-specific code for graph manipulation, thereby enabling more intuitive network management through natural language interactions. In [125], LLMs are employed to enhance the scalability of network functions as traffic volume increases. The authors introduce a system that utilizes LLMs to perform code analysis and extract crucial information about software behavior, semantics, and system-level performance. The extracted information is then used to optimize the infrastructure, deployment configuration, and execution pipeline. Shen et al. [121] propose to integrate LLMs into an Artificial Intelligence (AI)-enabled network with two distinct tasks: (i) serving as a user interface to intercept and understand user requests; and (ii) automating the training of AI nodes in the network—for instance, iteratively finding the best learning rate scheduler. In [117] and [119], LLMs are proposed as tools to facilitate the creation of network configurations from natural language descriptions. Both studies highlight the limitations of LLMs in this specific task, highlighting a high number of errors in the generated configurations. Consequently, both papers incorporate a verification module to validate and correct the output of the model. Ayed et al. [127] proposes a framework

<sup>7</sup><https://attack.mitre.org/tactics/ics/>

Table VII  
WORKS DEALING WITH NDA THROUGH GENAI MODELS (IN CHRONOLOGICAL ORDER).

Paper	Year	Input Type	GenAI Model				Evaluation/Fine-Tuning Datasets	Downstream Task	
			Name	Architecture	Fine-Tuned	Released		Target	Description
Soman and HG [116]	2023	Textual prompts	–	GPT-4, GPT-3.5, Bard, OpenAssistant-LLaMa	⊗	⊙	Cradlepoint†	👤	Digital assistants for telecom domain as user support
Wang et al. [117]	2023	Textual prompts	NetBuddy	GPT-4	⊗	⊙	self-built†	📱	Automate network configuration by translating requirements in natural language in low-level network configurations
Mani et al. [118]	2023	Graphs and textual prompts	–	GPT-4, GPT-3, GPT-3.5, Bard	⊗	⊙	NeMoEval	📱	Create task-specific code for graph analysis and manipulation
Mondal et al. [119]	2023	Textual prompts	–	GPT-4	⊗	⊙	self-built†	📱	Automate router configuration by translating prompts in natural language in router configurations
Roychowdhury et al. [120]	2024	Textual prompts	–	LLaMa 2.0	⊙	⊗	TeleQnA	👤	Digital assistants for question answering standard documents
Shen et al. [121]	2024	Textual prompts	–	GPT-3, GPT-4	⊗	⊙	self-built†	📱	Coordination of existing edge AI models to cater to the user's needs and enables automatic AI training
Duclos et al. [122]	2024	Text documents	–	CodeLLaMa	⊙	⊗	self-built†	📱	Translation of protocol specifications into structured models suitable for Cryptographic Protocol Shapes Analyzer
Ahmed et al. [123]	2024	Textual prompts	–	LLaMa 2.0, Falcon, Mistral 7B, Zephyr 7B-β	⊗	⊙	SPEC5GClassification, SPEC5GSummarization, TeleQnA	👤	Digital assistant for text classification, summarization, question answering
Piovesan et al. [124]	2024	Textual prompts	–	Phi-2, GPT-3.5, GPT-4	⊗	⊙	TeleQnA	👤	Digital assistants for question answering standard documents
Karapantelakis et al. [56]	2024	Textual prompts	–	GPT-3.5 turbo, GPT-4, LLaMa 2.0, Falcon, TeleRoBERTa	⊙	⊗	TeleQuAD†	👤	Digital assistants for Third Generation Partnership Projects
Ghasemirahni et al. [125]	2024	Code and textual prompts	FlowMage	GPT-3.5 turbo, GPT-4o, CodeLLaMa, Gemini 1.0 Pro	⊗	⊙	self-built†	📱	Software infrastructure optimization, deployment configuration, and execution pipeline
Erak et al. [126]	2024	Textual prompts	–	Phi-2, GPT-4o mini, GPT-4o	⊙	⊗	TeleQnA	👤	Digital assistants for Third Generation Partnership Projects
Ayed et al. [127]	2024	Textual prompts	Hermes	GPT-4o, LLaMa 3.1	⊗	⊙	self-built†	📱	Create logical blocks accompanied by code to execute specific networking intents

Fine-Tuned/Released GenAI Model Architecture: ⊙ - Present, ⊗ - Absent; Evaluation/Fine-Tuning Datasets: † - Private Data; Downstream Task: 👤 - The Output of the Model is Intended for a Human, 📱 - The Output of the Model is Intended for a Device.

based on LLMs to generate logical blocks related to a specific user intent (e.g., the deployment of a new base station in a network). Each logical block is accompanied by the corresponding code necessary for its implementation.

The “**Input Type**” column displays the specific data fed to GenAI models. Predominantly, these inputs are textual prompts where a query is submitted to the model. The majority of works employing such inputs fall under the category of downstream tasks for telecommunication documentation and support functions [56, 116, 120–124, 126]. Concerning the second downstream task—i.e., GenAI as assistant for the network operative phase—inputs vary depending on the specific operation. Specifically, Mani et al. [118] combine textual descriptions with network topology graphs, while Ghasemirahni et al. [125] integrate code snippets with textual prompts for analysis purposes. In [117, 119] a description of the configurations in natural language is used as input. Similarly, Ayed et al. [127] use a description of the available data together with a network modeling task.

Looking at the “**Evaluation/Fine-Tuning Datasets**” column, the datasets utilized in the literature vary according to the specific downstream task. TeleQnA is the most frequently used dataset [120, 124, 126, 128]. This dataset is designed to evaluate the knowledge of LLMs within the Telecom domain, featuring multiple-choice questions categorized into various categories. TeleQnAD [56] is similar to TeleQnA but special-

ized in 3GPP standards. On the other hand, NeMoEval [118] is used as benchmark for LLMs for two different applications: traffic analysis using communication graphs and network lifecycle management (e.g., capacity planning, network topology design, deployment planning, and diagnostic operations).

The “**GenAI Model**” column highlights that most studies, except for a few [117, 119, 120, 122], conduct comparative analyses among various currently-available generative models. The literature considers several LLMs, such as those from OpenAI (e.g., GPT-4o, GPT-4o mini, GPT-4, GPT-3.5, GPT-3), Google (e.g., Bard, Gemini), Meta (e.g., CodeLLaMa, LLaMa 2.0, LLaMa 3.1), and Microsoft (e.g., Phi-2). Regarding the use of open models, Mani et al. [118] test open LLMs (i.e., StarCoder and InCoder) but omitted their results due to inconsistency. Delving deeper, the majority of works [116–119, 121, 123–125, 127] do not fine-tune the models but use them off-the-shelf, as indicated by the “**Fine-Tuned**” column. On the other hand, the remaining works [56, 120, 122, 126] refine the models using domain-specific datasets. The “**Released**” column shows that none of the works that perform fine-tuning also release the fine-tuned models; only the original models are available in those cases. To reduce the computational burden of LLMs, the authors of [124, 126] investigate the use of Small Language Models (SLMs) leveraging Phi-2. In both works, such a model is compared with larger ones (i.e., GPT-3.5 and GPT-4, GPT-4o, GPT-4o mini). In [124],

the authors propose equipping the model with Retrieval-Augmented Generation (RAG) to incorporate authoritative knowledge external to the model's initial training data. The combination of Phi-2 and RAG achieves results comparable to those of GPT-3.5. Similarly, in [126], a fine-tuned Phi-2 in tandem with RAG and *SelfExtend* [129] (used to extend the model's the context window during inference) surpasses the performance achieved by the larger GPT-4o. Wang et al. [117] and Mondal et al. [119] introduce a verification module to address the shortcomings of GPT-4's device configuration generation. Their findings indicate that the LLM output frequently contains errors, necessitating a verification component to ensure accuracy and provide corrective feedback.

## V. A MODEL-CENTRIC OVERVIEW ON GENAI FOR NETWORK MONITORING AND MANAGEMENT

In this section, we present a model-centric overview of the works categorized in the previous sections. This section is divided into two parts: (i) the first part reports details about the base GenAI architecture leveraged by each work (Sec. V-A), while (ii) the second part focuses on modifications to GenAI architecture the authors performed in their proposals (Sec. V-B).

### A. Overview of GenAI Models

This section provides a broad view of the use of (foundation) GenAI models in NMM. Accordingly, Tab. VIII is centered around each “**GenAI Architecture**” and its application in the considered “**NMM Use Cases**”. First, we want to emphasize the persistent trend towards increasingly complex GenAI models. However, such complexity is not fully justified when considering the effectiveness of simpler ML/DL models in accomplishing the NMM tasks discussed in the present survey [131].

The GenAI architectures leveraged for network tasks defined in Sec. IV-A can be broadly divided—according to the nature of the underlying layers—into 5 categories (cf. Sec. III), namely: (i) Full Encoder-Decoder (FED), commonly leveraged for tasks where both the input and output are sequences, such as sequence-to-sequence tasks; (ii) Encoder-Only (EO), intended for language comprehension, specifically for interpreting and encoding input text for various subsequent applications; (iii) Decoder-Only (DO), frequently used for autoregressive text generation, meaning it generates text tokens based on the preceding token; (iv) Sequential Denoising Process (SDP), applied for producing high-quality data by means of successive denoising; (v) Selective and Structured State Space Model (SSM), representing a sophisticated means created for effectively handling and modeling long sequential data.

GenAI architectures like Transformer, T5, Gemini, Mistral, Zephyr, and XLNet<sup>8</sup> belong to the FED category. The EO category encompasses BERT along with its enhancements (e.g., DistilRoBERTa) and variations (e.g., CodeBERT), as well as ViT. The DO category features GPT, Falcon,

LLaMA, and Phi. The SDP category includes Diffusion and Stable Diffusion architectures, while the SSM category consists of the sole Mamba architecture.

Table VIII clearly shows that certain categories are more frequently utilized than others (column “**Cat.**”). The DO category accounts for the majority of works, followed by the FED and EO categories. When examining the models employed (“**Name**”), Transformer and BERT (by Google), the GPT family (by OpenAI), and LLaMA (by Meta) are the most widely used. Equally important, the reviewed works use most of the architectures designed for language-processing tasks. Accordingly, a large portion of them use plain-text-arranged information as model input (“**Input**”). Occasionally, this input is formatted as code, especially when the model is fine-tuned for code-generation tasks, like CodeBERT, CodeLLaMA, or CodeT5. Only a small fraction of the works [53, 75, 93] considers input traffic shaped as images.

A different perspective on GenAI models utilized for NMM use cases focuses on the research organization (“**Res. Org.**”) that introduced them and the associated licensing framework (“**Lic.**”). From this viewpoint, *three key points* emerge:

- *Non-academic organizations dominate the development of these architectures.* Google has been prolific, especially in the FED and EO categories, OpenAI has primarily developed DO solutions, and Meta and Microsoft have contributed to both the EO and DO GenAI categories.
- *The open-source paradigm is also embraced by non-academic entities.* Google, except for its private Bard and Gemini models, and Microsoft have released many models as open-source. In contrast, OpenAI's recent products, from GPT-3 onward, are closed-source. Meta, except for the open-source RoBERTa model, and StabilityAI employ non-commercial licenses.
- *Both academic and non-academic entities have pioneered each category of models.* These progenitor architectures are reported with a “†” in Tab. VIII. Google developed both Transformer and BERT models (the first in collaboration with the University of Toronto). OpenAI designed GPT-1. Diffusion and Mamba models are completely proposed by academic entities, namely, the former from Stanford and California universities and the latter from Carnegie Mellon and Princeton universities.

Concerning the specific use cases of the models in NMM (“**NMM Use Case**” column), BERT and GPT-like models address a wide range of applications, with BERT predominantly used for NID and GPT-like models for NDA. The Transformer model is applied to NTG, NTC, and NID. The LLaMA family has been exclusively used for NDA. Notably, the NTG use case is addressed by various GenAI model categories, including FED models like ViT, DO models such as GPT-2 and GPT-3, and diffusion-based models.

### B. Ad-Hoc GenAI Solutions

Table IX outlines the naming conventions used by state-of-the-art solutions. It specifies the base architecture and whether it is used “as-is” or fine-tuned, details the modifications made (if any), lists the components included in the pipeline

<sup>8</sup>XLNet, proposed by Google and Carnegie Mellon University in [130], combines a Transformer-based model with an AutoRegressive component.

Table VIII  
GENAI MODELS LEVERAGED IN THE NMM USE CASES CONSIDERED. GENAI MODELS ARE GROUPED BY CATEGORY AND ORDERED BY RELEASE YEAR WITHIN EACH CATEGORY.

GenAI Architecture						NMM Use Case				
Cat.	Name	Input	Year	Res. Org.	Lic.	NTG	NTC	NID	NSLA	NDA
FED	Transformer ‡	A	2017		●	[79]	[79][85]	[85][91][94][98][99]	–	–
	XLNet *	A	2019		●	–	–	–	[103]	–
	T5	A	2019		●	–	–	–	[110]	–
	CodeT5	</>	2020		●	–	–	–	[111]	–
	T5 1.1	A	2021		●	[78]	[78]	[78]	–	–
	Bard	A	2023		○	–	–	–	–	[116][118]
	Gemini 1.0 Pro	A	2023		○	–	–	–	–	[125]
	Mistral 7B	A	2023		●	–	–	–	–	[123]
	Zephyr 7B-β	A	2023		●	–	–	–	–	[123]
EO	BERT ‡	A	2018		●	–	[61][87]	[61][64][65][90][92][95][96]	[68][103][110][112][114]	–
	BERTiny	A	2019		●	–	–	–	[110]	–
	DistilRoBERTa	A	2019		●	–	–	–	[112]	–
	RoBERTa	A	2019		●	–	[86]	–	[112][113]	–
	BERT-of-Theseus	A	2020		●	–	–	[100]	–	–
	CodeBERT	</>	2020		●	–	–	–	[68]	–
	CodeBERTa	</>	2020		●	–	–	–	[68]	–
	ALBERT	A	2020		●	–	[88]	–	–	–
DO	ViT	🖼️	2021		●	–	–	[93]	–	–
	GPT-1 ‡	A	2018		●	–	–	[89][101]	[105][113]	–
	GPT-2	A	2019		●	[57]	[76]	[57]	[67][102][103][107][112]	–
	GPT-3	A	2020		○	[77]	–	[64]	[68][110]	[118][121]
	GPT-Neo	A	2021		○	–	–	–	[112]	–
	GPT-3.5	A	2022		○	–	–	–	[104]	[116][118][124]
	GPT-3.5 turbo	A	2022		○	–	–	[97]	[106][108][111]	[56][125]
	GPT-4	A	2023		○	–	–	–	–	[56][116][117][118][119][121][124]
	Falcon	A	2023		●	–	–	–	–	[56]
	OASST LLaMA	A	2023		○	–	–	–	–	[116]
	LLaMA 2.0	A	2023		○	–	–	–	–	[56][120][123]
	CodeLLaMA	</>	2023		○	–	–	–	–	[122][125]
	Phi 2.0	A	2023		●	–	–	–	–	[124][126]
	GPT-4o mini	A	2024		○	–	–	–	–	[126]
	GPT-4o	A	2024		○	–	–	–	–	[125][126][127]
	LLaMa 3.1	A	2024		○	–	–	–	–	[127]
SDP	Diffusion Model ‡	🖼️	2020		●	[75]	–	–	–	–
	Stable Diffusion 1.5	🖼️	2022		○	[53]	–	–	–	–
SSM	Mamba ‡	A	2023		●	[80]	[62]	[62]	–	–

#### LEGEND

**GenAI Model:** Category (Cat.), Research Organization (Res. Org.), Licensing (Lic.).

**NMM Use Case:** Network Traffic Generation (NTG), Network Traffic Classification (NTC), Network Intrusion Detection (NID), Networked System Log Analysis (NSLA), Network Digital Assistance (NDA).

**Category:** Fully Encoder-Decoder (FED), Encoder-Only (EO), Decoder-Only (DO), Sequential Denoising Process (SDP), State-Space Model (SSM).

**Input:** Natural Language (A), Code (</>), Image (🖼️).

**Research Organization:** Beihang University (🇨🇳), Carnegie Mellon University (🇺🇸), Deepmind (🇬🇧), EleutherAI (🇺🇸), Google Research (🇬🇪), Meta (🇺🇸), Microsoft (🇺🇸), MistralAI (🇫🇷), OpenAI (🇺🇸), OpenAssistant (🇺🇸), Princeton University (🇺🇸), Salesforce (🇺🇸), StabilityAI (🇺🇸), Stanford University (🇺🇸), Technological Innovation Institute (🇳🇱), Toyota Technological Institute at Chicago (🇯🇵), University of California Berkeley (🇺🇸), University of California San Diego (🇺🇸), University of Toronto (🇨🇦).

**Licensing:** open-source (●), non-commercial (○), proprietary (◐).

‡ The first row of each category contains the progenitor architecture.

\* XLNet is based on a Transformer architecture integrated with an AutoRegressive component [130].

before and after the GenAI model (if any), and indicates the availability of repositories for each proposed framework.

On the one hand, the majority of the works reported in Tab. IX (21 out of 28) utilize the base architecture in its vanilla version (see “V” column), typically adding pre-GenAI components like feature selectors/extractors and traffic-to-image modules, or post-GenAI components such as output refinement modules, exemplified by the ControlNet used in NetDiffusion. Among these, 8 works simply leverage the GenAI model without modifications or fine-tuning (see “F” column) but adding at least one pre-GenAI or post-GenAI component. Conversely, only 8 works propose modifications to the GenAI model and also perform fine-tuning for the targeted use cases.

For the NTG use case, notable examples include TrafficGPT [79] and LENS [78], both of which modify the

tokenizer to handle network traffic. LENS also changes the pre-training and fine-tuning phases. In the NTC and NID use cases, LENS [78] is again notable, along with netFound [85], which uses a Hierarchical Attention Transformer architecture, and ET-BERT [61], which redefines the tokenizer and the pre-training/fine-tuning procedures. Notably, none of the works addressing NSLA and NDA modify and fine-tune the base GenAI architecture, as these use cases align closely with the core philosophy of GenAI, e.g., document summarization and question-answering.

Finally, we also note whether a paper provides access to a related public repository (“Code Repo” column). Only 6 works make their framework code available [26, 53, 61, 62, 64, 67]. This lack of shared code significantly hampers reproducibility and hinders further development and verification by the research community.

Table IX

NAMING ADOPTED, BASE GENAI ARCHITECTURE, AND REPOSITORIES BY REVISED WORKS. WHEN THE PROPOSAL IS A TOOL/Framework, WE INDICATE ITS NAME WITH A ★ AND ONLY REPORT THE BEST-PERFORMING BASE GENAI ARCHITECTURE.

Base GenAI Architecture	Proposal Name	V	F	Modifications	Components		NMM Use Case					Paper	Year	Code Repo
					Pre GenAI	Post GenAI	NTG	NTC	NID	NSLA	NDA			
Transformer	ESeT	✓	✓	–	Multi-level Feature Extractor	Credibility Selector, Feature Augmentor	○	○	●	○	○	[91]	2022	–
	RTIDS	✓	✓	–	SMOTE, Feature Selection	–	○	○	●	○	○	[94]	2022	–
	netFound	✓	✓	Hierarchical Attention Transformer	–	–	○	●	●	○	○	[85]	2023	–
	TNN-IDS	✓	✓	–	Feature Extractor	–	○	○	●	○	○	[98]	2023	–
	RUIDS	✓	✓	Contrastive Loss	Sampling and Masking Module	Masked Context Reconstructor	○	○	●	○	○	[99]	2023	–
	TrafficGPT	✓	✓	Tokenizer	–	–	●	●	○	○	○	[79]	2024	–
TS 1.1	LENS	✓	✓	Tokenizer, Pre-Training and Fine-Tuning Stages	–	–	●	●	●	○	○	[78]	2024	–
BERT	ET-BERT	✓	✓	Tokenizer, Pre-Training and Fine-Tuning Stages	–	–	○	●	●	○	○	[61]	2022	🔗
	LAMBERT	✓	✓	–	Pre-Processing, Neighbor Sampling	–	○	●	○	○	○	[87]	2024	🔗
	Dom-BERT	✓	✓	–	Pre-Processing	–	○	○	○	●	○	[114]	2024	–
	SecurityBERT	✓	✓	Tokenizer	–	–	○	○	●	○	○	[65]	2024	–
BERT-of-Theseus	BT-TPF	✓	✓	KD Loss	Siamese Network	–	○	○	●	○	○	[100]	2024	–
CodeBERT	LogPrecis ★	✓	✓	–	–	–	○	○	○	●	○	[68]	2024	–
GPT-1	LogGPT	✓	✓	–	Log parser, Prompt Constructor	Response Parser	○	○	○	●	○	[105]	2023	–
GPT-2	GPT-2C	✓	✓	–	–	–	○	○	○	●	○	[102]	2021	–
	LogGPT	✓	✓	–	Prompt Generation	Reward Function	○	○	○	●	○	[67]	2023	🔗
	NetGPT	✓	✓	–	–	–	●	●	●	○	○	[57]	2023	–
GPT-3	PAC-GPT	✓	✓	–	–	Scapy Generator	●	○	○	○	○	[77]	2023	–
	FlowTransformer ★	✓	✓	–	Input Encoder	Classifier	○	○	●	○	○	[64]	2024	🔗
GPT-3.5	RAGLog	✓	✓	–	Log Database, Retriever	–	○	○	○	●	○	[104]	2023	–
GPT-3.5-turbo	HuntGPT	✓	✓	–	Explainer Module	–	○	○	●	○	○	[97]	2023	–
	LILAC	✓	✓	–	Adaptive Parsing Cache, ICL-Enhanced Parser	–	○	○	○	●	○	[106]	2023	–
	CYAGENT ★	✓	✓	–	Prompt Constructor, Prompt Generator	–	○	○	○	●	○	[111]	2024	–
GPT-4	NetBuddy	✓	✓	–	–	Output Verifier	○	○	○	○	●	[117]	2023	–
GPT-4o	FlowMage ★	✓	✓	–	Prompt Generator	–	○	○	○	○	●	[125]	2024	–
	Hermes ★	✓	✓	Chain of LLMs	–	–	○	○	○	○	●	[127]	2024	–
Diffusion Model	NetDiffus	✓	✓	–	Txt2Img Module	–	●	○	○	○	○	[75]	2023	–
Stable Diffusion 1.5	NetDiffusion	✓	✓	–	Txt2Img Module	ControlNet	●	○	○	○	○	[53]	2024	🔗
Mamba	NetMamba	✓	✓	–	Traffic Representation Module	–	○	●	●	○	○	[62]	2024	🔗

V: Vanilla Version; F: Fine-Tuned Version. **Architecture Category:** Fully Encoder-Decoder (FED), Encoder-Only (EO), Decoder-Only (DO), Sequential Denoising Process (SDP), State-Space Model (SSM). **NMM Use Case:** Network Traffic Generation (NTG), Network Traffic Classification (NTC), Network Intrusion Detection (NID), Networked System Log Analysis (NSLA), Network Digital Assistance (NDA).

## VI. DATASETS AND PLATFORMS

In the ever-changing networking landscape, GenAI emerges as a revolutionary tool that fosters innovative solutions and new perspectives. In this section, we explore two principal aspects of the GenAI tools that are used from a networking perspective, inspecting (i) the leveraged datasets for pre-training or fine-tuning/evaluation phases (Sec. VI-A) and (ii) the available platforms (Sec. VI-B) that facilitate the development of GenAI solutions for NMM.

### A. Datasets

Data play a critical role in applying GenAI in networking since they directly affect the development of AI models and their performance. In this section, we discuss the adoption of various datasets in the “Pre-Training”, “Fine-Tuning”, and “Evaluation” phases of a GenAI model life-cycle, when dealing with the NMM use cases defined in Sec. IV-A. Table X presents such an overview by associating *different colors* to the NMM use cases, along with general information about the datasets, namely their “Year” (according to the collection time-span) and the “Network Data” each dataset provides.<sup>9</sup>

<sup>9</sup>As presented in Sec. IV, different works could apply different pre-processing operations and extract different information from considered datasets to feed GenAI models.

It is worth noticing that the vast amount of unlabeled data collected (and now available) thanks to programmable devices and Software-Defined Networking (SDN) can be effectively utilized to refine the pre-training process of GenAI models, similarly as in other domains [166–168].

Still, we remark that many researchers do not release the data they used in their work, making it difficult to find publicly available datasets that are adequately representative of the specific context. To contribute valuable resources to the community and foster reproducibility, we focus on datasets that are *publicly available* or obtainable *upon-request* (35 out of 55). Notably, NDA is the NMM use case with fewer datasets reported in Tab. X; this is because most of them (i.e., 7 out of 10) are closed source (cf. Sec. IV-F).

Looking at the temporal evolution, older datasets tend to be more general and not explicitly collected or formatted for GenAI purposes. Conversely, more recent datasets [118, 154, 163, 164] are designed to optimize the training or evaluation of GenAI models. These datasets often include explicit prompts to better facilitate the generation of coherent and contextually relevant content. This is particularly true for NSLA and NDA use cases—being more affine to the NLP domain—where tailored datasets are used. In contrast, the other NMM use cases often repurpose traffic datasets originally collected for different objectives to train/evaluate GenAI models. For exam-

Table X  
OVERVIEW OF DATASETS USED IN THE PRE-TRAINING, TUNE-TUNING, AND EVALUATION PHASES OF GENAI MODELS IN THE NMM USE CASES CONSIDERED. DATASETS ARE REPORTED IN CHRONOLOGICAL ORDER.

Dataset Name	Year	Network Data	Pre-Training	Fine-Tuning	Evaluation
KDD'99 [132]	1999	Per-flow traffic features	–	[97][99]	[97][99]
ECML/PKDD 2007 [133]	2007	Web server logs	–	[112]	[112]
ISCX NSL-KDD [132]	2009	Per-flow traffic features	–	[64][96]	[64][96]
CSIC 2010 [134]	2010	HTTP requests	–	[92]	[92]
HttpParams [135]	2015	HTTP requests	–	[92]	[92]
UNSW-NB15 [136]	2015	Raw traffic data Per-flow traffic features	[76] –	[76] [64][93][99]	[76] [64][93][99]
IoT Sentinel [137]	2016	Traffic logs	–	–	[113]
ISCXVPN2016 [138]	2016	Raw traffic data Per-flow traffic features	[57][76][78][79] [57][61][62][78][79][87]	[57][76][78][79] [57][61][62][78][79][87]	[57][76][78][79] [57][61][62][78][79][87]
ISCTXor2016 [139]	2016	Raw traffic data Per-flow traffic features	[78][79] [78][79][62]	[78][79] [78][61][62]	[78][79] [78][61][62]
UNSW-IoT-Analytics [140]	2016	Raw traffic data & logs Per-flow traffic features	[76] –	[76] –	[76] [113]
USTC-TFC2016 [141]	2016	Raw traffic data	[57][79][78] [57][62][78][79] [57][62][78]	[57][78][79] [79] [57][61][62][78]	[57][78][79] [79] [57][61][62][78]
CIC-IDS2017 [142]	2017	Raw traffic data Per-flow traffic features	[61][87] [100]	– [64][85][91][93][94][99][100]	– [64][85][91][93][94][99][100]
Deep Fingerprinting [143]	2017	Per-flow traffic features	[75]	[75]	[75]
LabeledFlows [144]	2017	Per-flow traffic features	[81]	-	[81]
Cross-Platform [145]	2018	Raw traffic data	[79] [62]	[79] [61][62][79]	[79] [61][62][79]
CSE-CIC-IDS2018 [142]	2018	Raw traffic data Per-flow traffic features	[87] –	– [64][91]	– [64][91]
NLP2Bash [146]	2018	Unix shell logs	–	[68]	[68]
CIC-DDoS2019 [147]	2019	Per-flow traffic features	–	[94]	[94]
IoTFinder [148]	2019	Traffic logs	–	–	[113]
Web Server Access Logs [149]	2019	Web server logs & prompts	–	[111]	[111]
LabeledFlows [150]	2019	Per-flow traffic features	[81]	-	[81]
Apache Web Server [151]	2020	Web server logs	–	[112]	[112]
CIRA-CIC-DoHBrw2020 [152]	2020	Raw traffic data	[57][78][79] [57][78][79]	[57][78][79] [78]	[57][78][79] [78]
CyberLab Honeynet [153]	2020	Traffic logs	–	[68][102]	[68][102]
Loghub [154]	2020	System logs & prompts	–	[67][112]	[67][103][104][105][106][108][112]
MUDIS [155]	2020	Traffic logs	–	–	[113]
MQTT-IoT-IDS2020 [156]	2020	Raw traffic data Per-flow traffic features	[98]	[64][98]	[64][98]
TON_IoT [157]	2020	Raw traffic data Per-flow traffic features	– –	[77] [64][95][100]	[77] [64][95][100]
CSTNET-TLS1.3 [61]	2021	Raw traffic data	[87]	[61][87]	[61][87]
FWAF [158]	2021	HTTP requests	–	[92]	[92]
Spirit [159]	2021	System logs	–	[105][112]	[112]
CIC-IoT-2022 [160]	2022	Raw traffic data Per-flow traffic features	[79] [79][62] [62]	[79] – [62]	[79] – [62]
Edge-IIoTset [161]	2022	Raw traffic data Per-flow traffic features	–	[65][101]	[65][101]
CIC-IoT-2023 [162]	2023	Raw traffic data Per-flow traffic features	[78] [78]	– –	– –
NeMoEval [118]	2023	Network graphs & prompts	–	–	[118]
SPEC5G [163]	2023	Prompts	–	–	[123]
TeleQnA [164]	2023	Prompts	–	[120][123][124][126]	[120][123][124][126]
Censys [165]	2024	Traffic logs HTTP requests	[86] –	[86] –	[86] [113]

Use Cases: Network Traffic Generation (NTG) – Network Traffic Classification (NTC) – Network Intrusion Detection (NID) – Networked System Log Analysis (NSLA) – Network Digital Assistance (NDA) .  
The datasets highlighted with more than one color are used for multiple use cases.

ple, in [61] numerous raw-traffic datasets collected in different domains [61, 139, 142, 145] are fused to fine-tune the ET-BERT model and allow it to deal with multiple NTC/NID-related tasks (e.g., NTC on VPN/TLS/Tor, malware classification).

Other trends can be inferred when considering the phases of a GenAI model life-cycle. On the one hand, the datasets mostly used for pre-training are typically more generic than those used in the other phases, since they aim to allow the model learning a foundational understanding of the context. On the other hand, the datasets used for fine-tuning are vertical to the specific NMM use case, as their goal is to specialize the pre-trained model. Similarly, the evaluation phase requires data specific to the NMM use case to be solved. Therefore, most fine-tuning datasets are also used for evaluation. Additionally, the evaluation datasets are usually meticulously labeled to facilitate accurate quantitative assessment of the related models in a supervised manner. Unfortunately, regarding this latter aspect, different methods often rely on distinct datasets for performance evaluation. This inconsistency results in significant efforts for manual data processing and ultimately in unfair comparisons. To address this limitation, initial large-scale and unified benchmark datasets specifically designed for assessing GenAI models are beginning to be proposed [169]. These datasets—particularly for validating foundation models in NTC and NTG use cases—aim to replace the current reliance on heterogeneous compositions of older datasets.

From Tab. X, we can also extract some differences in how data are used for different NMM use cases. Interestingly, all the works falling within the NTC (colored in pale green) train from scratch the GenAI architecture on one or more datasets. On the other hand, as can be noted from the “pre-training” column, all the works that perform NSLA and NDA, along with the majority addressing intrusion detection, start with a pre-trained model and only fine-tune it. While it is expected for NSLA and NDA use cases, whose models are usually fed with textual inputs (and can thus leverage pre-trained LLMs), this does not apply to works performing intrusion detection. Indeed, the latter commonly employ models pre-trained for affine tasks, such as NTC (and vice versa). Accordingly, for NTC, some datasets [142, 162] are used exclusively for pre-training purposes. Finally, datasets employed in NTG are commonly leveraged for other tasks as well, except for NDA, which, as aforementioned, relies on ad-hoc GenAI-tailored data.

Table X highlights the presence of a relatively large number of publicly available datasets. However, it is crucial for the scientific community to focus on the quality of these datasets. Many datasets fail to adequately represent real-world scenarios due to various factors, such as the use of synthetically generated data, small-scale testbeds, or heavy anonymization [170]. For instance, CIC-IDS2017, which is the most common dataset for the NID task (cf. Sec. IV-D), suffers from several issues that hinder its utility as a benchmark [171]. Therefore, an important step for the scientific community would be to critically filter public datasets to select only those that provide a reliable representation of real-world scenarios.

## B. Dedicated Platforms

GenAI solutions require huge amounts of resources to be effective—during both the pre-training and the fine-tuning phases. Thus the deployment of these solutions can benefit from simplifications in the management of computing infrastructures, which translates into reduced maintenance costs. This calls for *dedicated platforms* that offer integrated environments that streamline the end-to-end AI workflow. These platforms provide essential tools and services for data processing, model training, deployment, and monitoring.

In the context of GenAI, platforms can be categorized based on two main aspects: (a) whether they provide only first-party models (i.e., developed by the platform owner) or also include third-party models, and (b) whether they offer open-source models in addition to closed-source ones, which are typically accessible exclusively through User Interfaces (UIs), Application Programming Interfaces (APIs), or Command Line Interfaces (CLIs).

Here we survey some of the most well-known and utilized platforms according to this categorization.

To the best of our knowledge, no platform that offers only **first-party** models, provides them as **open source**. *OpenAI GPT* is a notable case that offers access to a vast amount of **first-party** pre-trained natural language processing models—such as GPT and its advancements—which can be fine-tuned using computational resources provided by OpenAI. However, the models are **closed**, and they remain available and can be operated only on OpenAI servers. The functionality of OpenAI GPT can be accessed through UI/API via free/paid plans or by signing a partnership.

On the other hand, most platforms offer both **first-party and third-party** models, including **open-source and closed** alternatives. In this case, the user leverages CLI, UI, or interactive notebooks for performing pre-training, fine-tuning, and operation of models. Notably, the business models of these platforms allow users to *deploy and execute even open-source models only on their own cloud* with different degrees of customization. In more detail, *Amazon Bedrock* is a cloud computing platform that provides access to several foundation models from various companies (e.g., AI21 Labs, Anthropic, Cohere, Meta, Mistral AI, Stability AI, and Amazon). Similarly, *Microsoft Azure AI* offers a wide range of services for GenAI, including pre-built and customizable APIs and models. The users can train models on the Azure cloud with their data and access a variety of pre-trained models (from OpenAI, Hugging Face, Stability AI, Meta, etc.). *Google Cloud AI Platform (Vertex AI)* provides services for the development and fine-tuning of pre-trained LLMs from Google (i.e., Gemini, Gemma) and other open models (e.g., LLaMa, Claude). Users can customize hardware resources (e.g., GPUs, storage, and virtual machines) according to their needs.

Other platforms provide only **open-source** models, both **first- and third-party**. *Hugging Face*, one of the most popular and active platforms in the AI and particularly GenAI community, belongs to this category. Hugging Face hosts a vast amount of open-source pre-trained models and also provides remote resources for model training via API on a subscription

basis. *Nvidia NIM* has a more specific focus on provider hardware support and exposes an API to access to numerous open pre-trained LLMs hosted on its infrastructure. These models are optimized for Nvidia architectures and accelerated through the Nvidia software stack. Differently than these platforms that allow the users to *download, customize, and operate the models outside of them*, *IBM Watson* provides pre-trained models (e.g., Granite, LLaMa, Mixtral) for GenAI deployed on its cloud, which can be accessed only via APIs and notebooks. Based on a slightly different philosophy, *Cloudflare AI* is based on a network of serverless GPUs specifically designed for deploying and running AI models from anywhere. Cloudflare AI offers numerous open models for text generation and text-to-image tasks (e.g., LLaMa, Gemma, Zephyr, StableDiffusion) whose interaction can be carried out via API and CLI.

Among the frameworks categorized in Tab. VIII (cf. Sec. V), all the base architectures are available on (and downloadable from) Hugging Face, except for the (closed) GPT-based models, which are available on OpenAI, and BERT, which is accessible through the Google Cloud AI Platform.

## VII. WRAP UP AND FUTURE PROSPECTS

The advent of GenAI presents a transformative potential for network monitoring and management tasks. By leveraging advanced generative models, network systems can address the growing complexity and dynamic nature of modern networks with a more predictive and proactive stance, moving beyond the traditional reactive measures.

Despite the positive aspects of this breakthrough approach, GenAI also has several *limitations* summarized in the left side of Figure 9. One of the major concerns is (i) *the lack of trustworthiness and robustness* [170]. The issue of trustworthiness is due to “closed box” nature of the GenAI-model architectures. Hence, users often find it difficult to understand the decision-making processes of a GenAI model, thus complicating the debugging, refinement, and efforts to improve model performance and address biases. Furthermore, the evaluation of GenAI models focuses only on a few publicly-available datasets, thus not reflecting dynamic real-world scenarios. Additionally, network attacks or crafted inputs to evade malware identification can *intentionally* modify the nature of network traffic. These points raise doubts about the robustness of GenAI solutions and their applicability to network monitoring and management in practice.

Furthermore, (ii) *these models require substantial computational resources for pre-training and fine-tuning*. Achieving high performance typically involves processing extensive datasets, which demands significant computational power and time. Additionally, current GenAI models consist of trillions of parameters and require extensive training periods and significant resources that may not be available in smaller computational environments. In fact, some studies [124, 126] suggest using more compact models (e.g., SLM) for network monitoring and management tasks, which, with a significantly smaller number of parameters, can deliver performance comparable to larger models. This high resource requirement represents a major hindrance to the widespread adoption of GenAI

and poses challenges for its online application in network environments. As a result, integrating such models into operational networks remains complex and costly. Additionally, the complexity of these models also introduces (iii) *potential vulnerabilities to adversarial attacks*, posing security risks. Attackers can exploit their architectures, manipulating their behavior or compromising data integrity. Expressly, adversarial attacks may use subtle perturbations to trick the model into making incorrect predictions, risking data security and system reliability. Lastly, the use of GenAI raises (iv) *ethical concerns* related to data privacy and misuse of generated content. Personal data are used in training models, potentially causing significant privacy issues. Then, the ability to generate misleading content (such as deepfakes) can lead to misinformation and manipulation.

In the following, we identify possible **future directions** to improve and overcome the drawbacks of GenAI in network monitoring and management. We graphically summarize these perspectives on the right side of Figure 9.

**Real-time suitability and efficiency:** for network monitoring and management tasks, timely operation is a strict requirement. Nevertheless, GenAI models demand substantial computational power and extensive datasets for both training and fine-tuning. *Federated Learning* [172, 173] can help overcome these challenges by enabling decentralized training across multiple devices, thereby minimizing the need for centralized data storage and processing [174]. Additionally, large model sizes can be a significant bottleneck, resulting in longer inference times. A key future direction is to develop efficient GenAI models that can be effectively trained on smaller datasets. Furthermore, model compression, efficient architecture design, and hardware acceleration can improve the processing speed and reduce computational overhead. In this direction, integrating *Green AI* [175] principles in GenAI development allows an environmentally-sustainable progress, reducing the energy consumption and carbon footprint associated with model training and deployment. Specifically, *TinyML* [176, 177] can offer a powerful combination of efficiency and intelligence. TinyML enables real-time, low-power data processing on edge devices, while GenAI provides advanced predictive modeling and simulation. This synergy allows for proactive network management, offering localized insights and responses that enhance network efficiency, even in resource-constrained environments (e.g., when running GenAI models directly on handheld devices). Additionally, quick model adaptation to network shifts and anomalies is crucial for maintaining efficient and responsive network operations. In this context, *Quantum ML* [178, 179] may significantly boost this capability in the long term, speeding up model training and enhancing predictive precision in intricate and evolving scenarios.

**Handle network data complexity:** integrating *multimodal GenAI* promises to revolutionize network monitoring and management. By leveraging the capability of AI to process and analyze diverse data types—ranging from textual logs and metrics to visual network topologies—network operators can achieve unprecedented levels of insight and automation.

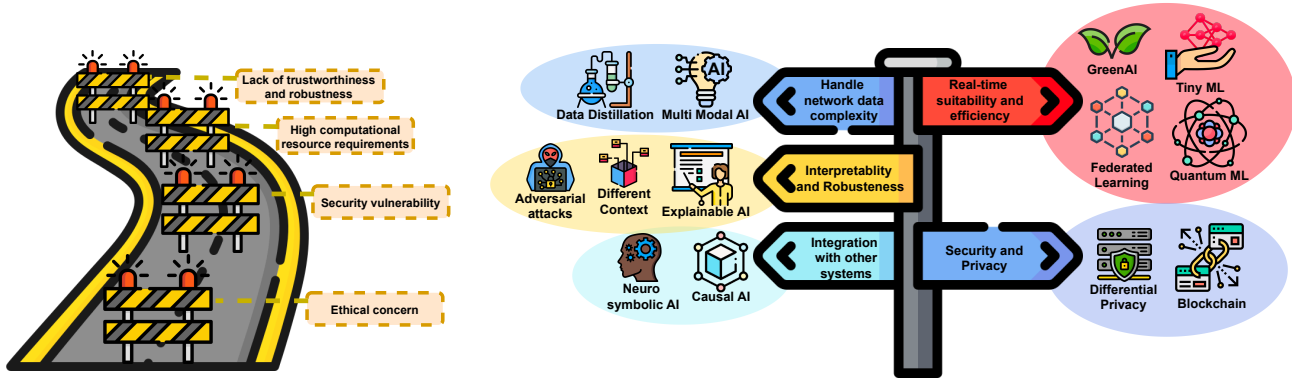


Figure 9. Limitations (*left-side*), represented by roadblocks that highlight the current challenges in the adoption and implementation of GenAI for network monitoring and management. Future directions (*right-side*), grouped into five categories distinguished by different colors, aimed at addressing these limitations.

This holistic approach allows for more accurate anomaly detection, predictive maintenance, and dynamic resource allocation, ultimately leading to more resilient and efficient network infrastructures. The ability of multimodal AI to synthesize information from multiple sources enhances “on-the-fly” decision-making while paving the way for adaptive, self-healing networks, aligning with the vision of a fully automated, intelligent network management paradigm. Furthermore, more advanced techniques, such as Reinforcement Learning [180, 181], enhance the adaptivity of LLMs, enabling them to evolve during their operational mode. Additionally, *data distillation* procedures can be enforced to manage the large scale and redundancy of datasets. This approach helps in reducing the dataset size by retaining only the most essential samples and discarding unnecessary ones. Such a procedure can significantly decrease both the training time and the resources required.

**Interpretability and robustness:** the convergence of XAI [182, 183] and GenAI heralds a new era in network monitoring and management, where transparency and innovation go hand in hand. XAI provides much-needed clarity in the decision-making processes of AI systems, enabling network operators to trust and understand the actions taken by their automated tools [184]. When coupled with the GenAI’s ability to simulate and predict network behaviors, this synergy offers a robust framework for proactive management. For instance, network anomalies can be not only detected but also explained leveraging XAI tools and addressed with AI-generated tailored responses. This fusion ensures actionable and transparent AI-driven insights, fostering a deeper integration of AI in network operations. As a result, network management becomes more intelligent, reliable, and user-centric, enabling networks to be both self-optimizing and comprehensible. Additionally, to improve model robustness, we can leverage diverse and comprehensive datasets to improve the *generalizability of GenAI models across different contexts*, optionally including multi-task learning techniques, and their *resistance to adversarial attacks*, exploiting adversarial training or data augmentation strategies.

**Integration with other systems:** integrating GenAI models with existing network systems enhances their utility in network monitoring and management by ensuring interoperability with the current infrastructure, developing robust APIs for seamless embedding, and enabling data fusion from multiple sources. This integration also ensures scalability and leverages automation to handle routine tasks, providing comprehensive and intelligent network management solutions. Future directions could focus on integrating *Causal AI* [185], which emphasizes understanding cause-and-effect relationships, and *Neuro-symbolic AI* [186, 187], which combines the learning capabilities of neural networks with the logical reasoning of symbolic AI into GenAI models. This integration could improve the ability of these models to handle complex long-term tasks and multi-step decision-making, empowering them to accomplish intelligent planning.

**Security and privacy:** as GenAI models are increasingly used in network monitoring and management, ensuring their security and privacy is essential. Future directions should focus on integrating *Blockchain* [188, 189] technology can provide a decentralized and immutable framework for secure data sharing and model updates, further enhancing the overall security and transparency of GenAI applications. Moreover, implementing privacy-preserving techniques, such as *differential privacy*, can safeguard user data, and ensure secure deployment with robust access control and secure communication channels. In addition, addressing ethical considerations by ensuring transparent data usage, unbiased model training, and accountability in AI decisions will be crucial in building trust and reliability in GenAI applications for network management.

In conclusion, integrating GenAI into network monitoring and management holds significant promise. However, it is crucial to carefully manage expectations and avoid overly optimistic assumptions about its capabilities. Tackling the associated limitations is essential to ensure a realistic and effective implementation. Collaboration between academia and industry is vital to ensure that the generative models developed are not only theoretically sound but also practical and scalable in real-world applications. Establishing benchmarks

and standardized datasets to evaluate the performance of GenAI in network monitoring and management can provide a foundation for continuous improvement and innovation. By driving advancements in predictive analytics, anomaly detection, and automation, future research can pave the way for more intelligent, efficient, and secure network systems. As we continue to explore this intersection of GenAI and networking, it is imperative to address the associated challenges and ethical considerations to fully harness the potential of GenAI.

#### ACKNOWLEDGMENTS

This work is partially supported by the European Union under the Italian National Recovery and Resilience Plan (NRRP) of NextGenerationEU, partnership on “Telecommunications of the Future” (PE00000001 – program “RESTART”). Also, this work is partially carried out within the “xInternet” Project supported by the MUR PRIN 2022 program (D.D.104—02/02/2022) funded by the NextGenerationEU.

#### REFERENCES

- [1] J. Yang, H. Jin, R. Tang, X. Han, Q. Feng, H. Jiang, S. Zhong, B. Yin, and X. Hu, “Harnessing the power of llms in practice: A survey on chatgpt and beyond,” *ACM Transactions on Knowledge Discovery from Data*, vol. 18, no. 6, pp. 1–32, 2024.
- [2] “Generative Artificial Intelligence (AI) Market Size Worldwide from 2020 to 2030,” <https://www.statista.com/forecasts/1449838/generative-ai-market-size-worldwide>.
- [3] X. Huang, H. Yang, C. Zhou, X. Shen, and W. Zhuang, “When Digital Twin Meets Generative AI: Intelligent Closed-Loop Network Management,” *arXiv preprint arXiv:2404.03025*, 2024.
- [4] H. Du, D. Niyato, J. Kang, Z. Xiong, P. Zhang, S. Cui, X. Shen, S. Mao, Z. Han, A. Jamalipour *et al.*, “The Age of Generative AI and AI-generated Everything,” *IEEE Network*, 2024.
- [5] A. Lavin, C. M. Gilligan-Lee, A. Visnjic, S. Ganju, D. Newman, S. Ganguly, D. Lange, A. G. Baydin, A. Sharma, A. Gibson *et al.*, “Technology readiness levels for machine learning systems,” *Nature Communications*, vol. 13, no. 1, p. 6039, 2022.
- [6] D. Rossi and L. Zhang, “Landing AI on Networks: An Equipment Vendor Viewpoint on Autonomous Driving Networks,” *IEEE Transactions on Network and Service Management*, vol. 19, no. 3, pp. 3670–3684, 2022.
- [7] “The Networking Channel,” <https://networkingchannel.eu/library/>.
- [8] “AT&T’s new Generative AI Tool Will Help Employees Be More Effective, Creative, and Innovative,” <https://about.att.com/blogs/2023/generative-ai.html>.
- [9] “Cisco Artificial Intelligence,” <https://www.cisco.com/site/us/en/solutions/artificial-intelligence/ai-assistant/index.html>.
- [10] “How to Make Better Use of Network Insights with Generative AI,” <https://www.ericsson.com/en/blog/2024/2/how-to-make-better-use-of-network-insights-with-generative-ai>.
- [11] “European Innovation Council,” [https://eic.ec.europa.eu/index\\_en](https://eic.ec.europa.eu/index_en).
- [12] “Huawei Introduces AI Technologies to Accelerate Network Transformation Towards All Intelligence in the Net5.5G Era,” <https://www.huawei.com/en/news/2024/4/has-net-5-point-5g-ai>.
- [13] “Welcome to the Large Generative AI Models in Telecom (GenAINet) Emerging TechnologyInitiative website,” <https://genainet.committees.comsoc.org/>.
- [14] “IETF Side Meetings,” <https://wiki.ietf.org/en/meeting/119/sidemeetings>.
- [15] “Specializing Large Language Models for Telecom Networks,” <https://aiforgood.itu.int/event/specializing-large-language-models-for-telecom-networks/>.
- [16] “Generative AI implications for Telco Operations,” <https://www.bell-labs.com/institute/white-papers/generative-ai-implications-for-telco-operations/>.
- [17] “Telefónica Partners with Microsoft to Incorporate Generative AI into Kernel,” <https://www.telefonica.com/en/communication-room/press-room/telefonica-partners-with-microsoft-to-incorporate-generative-ai-into-kernel/>.
- [18] “Generative AI: the challenge of TIM for the future of IT,” [https://www.gruppottim.it/it/newsroom/notiziario-tecnico-tim/Anno-2023/n3-2023/Generative\\_AI\\_la\\_sfida\\_TIM\\_per\\_il\\_futuro\\_dell\\_IT.html](https://www.gruppottim.it/it/newsroom/notiziario-tecnico-tim/Anno-2023/n3-2023/Generative_AI_la_sfida_TIM_per_il_futuro_dell_IT.html).
- [19] S. Sai, M. Kanadia, and V. Chamola, “Empowering IoT with Generative AI: Applications, Case Studies, and Limitations,” *IEEE Internet of Things Magazine*, vol. 7, no. 3, pp. 38–43, 2024.
- [20] M. Hassanin and N. Moustafa, “A Comprehensive Overview of Large Language Models (LLMs) for Cyber Defences: Opportunities and Directions,” *arXiv preprint arXiv:2405.14487*, 2024.
- [21] F. Alwahedi, A. Aldaheri, M. A. Ferrag, A. Battah, and N. Tihanyi, “Machine Learning Techniques for IoT Security: Current Research and Future Vision with Generative AI and Large Language Models,” *Internet of Things and Cyber-Physical Systems*, vol. 4, pp. 167–185, 2024.
- [22] J. Halvorsen, C. Izurieta, H. Cai, and A. Gebremedhin, “Applying Generative Machine Learning to Intrusion Detection: A Systematic Mapping Study and Review,” *ACM Computing Surveys*, vol. 56, no. 10, 2024.
- [23] H. Zhou, C. Hu, Y. Yuan, Y. Cui, Y. Jin, C. Chen, H. Wu, D. Yuan, L. Jiang, D. Wu *et al.*, “Large Language Model (LLM) for Telecommunications: A Comprehensive Survey on Principles, Key Techniques, and Opportunities,” *arXiv preprint arXiv:2405.10825*, 2024.
- [24] A. Celik and A. M. Eltawil, “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.
- [25] A. Karapantelakis, P. Alizadeh, A. Alabassi, K. Dey, and A. Nikou, “Generative AI in Mobile Networks: a Survey,” *Annals of Telecommunications*, vol. 79, no. 1, pp. 15–33, 2024.
- [26] C. Liu, X. Xie, X. Zhang, and Y. Cui, “Large Language Models for Networking: Workflow, Advances and Challenges,” *arXiv preprint arXiv:2404.12901*, 2024.
- [27] Y. Huang, H. Du, X. Zhang, D. Niyato, J. Kang, Z. Xiong, S. Wang, and T. Huang, “Large Language Models for Networking: Applications, Enabling Techniques, and Challenges,” *arXiv preprint arXiv:2311.17474*, 2023.
- [28] C. Chaccour, A. Karapantelakis, T. Murphy, and M. Dohler, “Telecom’s Artificial General Intelligence (AGI) Vision: Beyond the GenAI Frontier,” *IEEE Network*, 2024.
- [29] M. A. Ferrag, F. Alwahedi, A. Battah, B. Cherif, A. Mechri, and N. Tihanyi, “Generative AI and Large Language Models for Cyber Security: All Insights You Need,” *arXiv preprint arXiv:2405.12750*, 2024.
- [30] M. Xu, H. Du, D. Niyato, J. Kang, Z. Xiong, S. Mao, Z. Han, A. Jamalipour, D. I. Kim, X. Shen *et al.*, “Unleashing the Power of Edge-Cloud Generative AI in Mobile Networks: A Survey of AIGC Services,” *IEEE Communications Surveys & Tutorials*, 2024.
- [31] J. Wang, H. Du, D. Niyato, J. Kang, Z. Xiong, D. I. Kim, and K. B. Letaief, “Toward Scalable Generative AI via Mixture of Experts in Mobile Edge Networks,” *arXiv preprint arXiv:2402.06942*, 2024.
- [32] G. Aceto, G. Bovenzi, D. Ciuonzo, A. Montieri, V. Persico, and A. Pescapé, “Characterization and prediction of mobile-app traffic using Markov modeling,” *IEEE Transactions on Network and Service Management*, vol. 18, no. 1, pp. 907–925, 2021.
- [33] G. Aceto, F. Giampaolo, C. Guida, S. Izzo, A. Pescapé, F. Piccialli, and E. Prezioso, “Synthetic and Privacy-Preserving Traffic Trace Generation using Generative AI Models for Training Network Intrusion Detection Systems,” *Journal of Network and Computer Applications*, p. 103926, 2024.
- [34] S. Hui, H. Wang, Z. Wang, X. Yang, Z. Liu, D. Jin, and Y. Li, “Knowledge enhanced GAN for IoT traffic generation,” in *ACM Web Conference (WWW)*, 2022, pp. 3336–3346.
- [35] D. Gudovskiy, S. Ishizaka, and K. Kozuka, “CFLOW-AD: real-time unsupervised anomaly detection with localization via conditional normalizing flows,” in *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2022, pp. 98–107.
- [36] D. P. Kingma and M. Welling, “Auto-encoding variational Bayes,” in *International Conference on Learning Representations (ICLR)*, 2014.
- [37] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, “Attention is All You Need,” *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017.
- [38] L. Dinh, D. Krueger, and Y. Bengio, “NICE: Non-linear independent components estimation,” in *International Conference on Learning Representations (ICLR), Workshop Track*, 2015.
- [39] O. A. Adeleke, N. Bastin, and D. Gurkan, “Network Traffic Generation: A Survey and Methodology,” *ACM Computing Surveys*, vol. 55, no. 2, pp. 1–23, 2022.

- [40] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 27, 2014.
- [41] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, 2014.
- [42] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," in *International Conference on Learning Representations (ICLR)*, 2016.
- [43] A. B. L. Larsen, S. K. Sønderby, H. Larochelle, and O. Winther, "Autoencoding beyond pixels using a learned similarity metric," in *International Conference on Machine Learning (ICML)*, 2016, pp. 1558–1566.
- [44] L. Dinh, J. Sohl-Dickstein, and S. Bengio, "Density estimation using Real NVP," in *International Conference on Learning Representations (ICLR)*, 2017.
- [45] R. Bommasani, D. A. Hudson, E. Adeli, R. Altman, S. Arora, S. von Arx, M. S. Bernstein, J. Bohg, A. Bosselut, E. Brunskill *et al.*, "On the opportunities and risks of foundation models," *arXiv preprint arXiv:2108.07258*, 2021.
- [46] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *arXiv preprint arXiv:1810.04805*, 2018.
- [47] A. Radford, K. Narasimhan, T. Salimans, I. Sutskever *et al.*, "Improving Language Understanding by Generative Pre-Training," *OpenAI Tech Report*, 2018.
- [48] J. Ho, A. Jain, and P. Abbeel, "Denoising Diffusion Probabilistic Models," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 33, pp. 6840–6851, 2020.
- [49] D. Kingma and R. Gao, "Understanding diffusion objectives as the elbo with simple data augmentation," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 36, 2024.
- [50] A. Gu and T. Dao, "Mamba: Linear-Time Sequence Modeling with Selective State Spaces," *arXiv preprint arXiv:2312.00752*, 2023.
- [51] J. Kim, J. H. Lee, S. Kim, J. Park, K. M. Yoo, S. J. Kwon, and D. Lee, "Memory-efficient fine-tuning of compressed large language models via sub-4-bit integer quantization," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 36, 2024.
- [52] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, "LoRa: Low-rank adaptation of large language models," in *International Conference on Learning Representations (ICLR)*, 2022.
- [53] X. Jiang, S. Liu, A. Gember-Jacobson, A. N. Bhagoji, P. Schmitt, F. Bronzino, and N. Feamster, "NetDiffusion: Network Data Augmentation Through Protocol-Constrained Traffic Generation," *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, vol. 8, no. 1, pp. 1–32, 2024.
- [54] N. Ding, Y. Qin, G. Yang, F. Wei, Z. Yang, Y. Su, S. Hu, Y. Chen, C.-M. Chan, W. Chen *et al.*, "Parameter-efficient fine-tuning of large-scale pre-trained language models," *Nature Machine Intelligence*, vol. 5, no. 3, pp. 220–235, 2023.
- [55] "GGUF." [Online]. Available: <https://github.com/ggerganov/ggml/blob/master/docs/gguf.md>
- [56] A. Karapantelakis, M. Shakur, A. Nikou, F. Moradi, C. Orlog, F. Gaim, H. Holm, D. D. Nimara, and V. Huang, "Using Large Language Models to Understand Telecom Standards," in *IEEE International Conference on Machine Learning for Communication and Networking (ICMLCN)*, 2024.
- [57] X. Meng, C. Lin, Y. Wang, and Y. Zhang, "NetGPT: Generative Pretrained Transformer for Network Traffic," *arXiv preprint arXiv:2304.09513*, 2023.
- [58] F. Pacheco, E. Exposito, M. Gineste, C. Baudoin, and J. Aguilar, "Towards the Deployment of Machine Learning Solutions in Network Traffic Classification: A Systematic Survey," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 2, pp. 1988–2014, 2018.
- [59] A. Azab, M. Khasawneh, S. Alrabaa, K.-K. R. Choo, and M. Sarsour, "Network Traffic Classification: Techniques, Datasets, and Challenges," *Digital Communications and Networks*, 2022.
- [60] G. Aceto, D. Cionzo, A. Montieri, V. Persico, and A. Pescapè, "AI-powered Internet Traffic Classification: Past, Present, and Future," *IEEE Communications Magazine*, pp. 1–7, 2023.
- [61] X. Lin, G. Xiong, G. Gou, Z. Li, J. Shi, and J. Yu, "ET-BERT: A Contextualized Datagram Representation with Pre-training Transformers for Encrypted Traffic Classification," in *ACM Web Conference (WWW)*, 2022, p. 633–642.
- [62] T. Wang, X. Xie, W. Wang, C. Wang, Y. Zhao, and Y. Cui, "NetMamba: Efficient Network Traffic Classification via Pre-training Unidirectional Mamba," *arXiv preprint arXiv:2405.11449*, 2024.
- [63] D. Chou and M. Jiang, "A Survey on Data-driven Network Intrusion Detection," *ACM Computing Surveys*, vol. 54, no. 9, pp. 1–36, 2021.
- [64] L. D. Manocchio, S. Layeghy, W. W. Lo, G. K. Kulatilake, M. Sarhan, and M. Portmann, "FlowTransformer: A Transformer Framework for Flow-based Network Intrusion Detection Systems," *Expert Systems with Applications*, vol. 241, p. 122564, 2024.
- [65] M. A. Ferrag, M. Ndhlovu, N. Tihanyi, L. C. Cordeiro, M. Debbah, T. Lestable, and N. S. Thandi, "Revolutionizing Cyber Threat Detection With Large Language Models: A Privacy-Preserving BERT-based Lightweight Model for IoT/IIoT Devices," *IEEE Access*, vol. 12, pp. 23 733–23 750, 2024.
- [66] S. He, P. He, Z. Chen, T. Yang, Y. Su, and M. R. Lyu, "A Survey on Automated Log Analysis for Reliability Engineering," *ACM computing surveys*, vol. 54, no. 6, pp. 1–37, 2021.
- [67] X. Han, S. Yuan, and M. Trabelsi, "LogGPT: Log Anomaly Detection via GPT," in *IEEE International Conference on Big Data (BigData)*, 2023, pp. 1117–1122.
- [68] M. Boffa, I. Drago, M. Mellia, L. Vassio, D. Giordano, R. Valentim, and Z. B. Houidi, "LogPrécis: Unleashing Language Models for Automated Malicious Log Analysis: Précis: A Concise Summary of Essential Points, Statements, or Facts," *Computers & Security*, vol. 141, p. 103805, 2024.
- [69] A. Martinez, M. Yannuzzi, V. López, D. López, W. Ramírez, R. Serral-Gracià, X. Masip-Bruin, M. Maciejewski, and J. Altmann, "Network Management Challenges and Trends in Multi-Layer and Multi-Vendor Settings for Carrier-Grade Networks," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 4, pp. 2207–2230, 2014.
- [70] L. Aarikka-Stenroos and P. Ritala, "Network Management in the Era of Ecosystems: Systematic Review and Management Framework," *Industrial Marketing Management*, vol. 67, pp. 23–36, 2017.
- [71] M. Aboubakar, M. Kellil, and P. Roux, "A Review of IoT Network Management: Current Status and Perspectives," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 7, pp. 4163–4176, 2022.
- [72] J. Wang, L. Zhang, Y. Yang, Z. Zhuang, Q. Qi, H. Sun, L. Lu, J. Feng, and J. Liao, "Network Meets ChatGPT: Intent Autonomous Management, Control and Operation," *Journal of Communications and Information Networks*, vol. 8, no. 3, pp. 239–255, 2023.
- [73] J. Holland, P. Schmitt, N. Feamster, and P. Mittal, "New Directions in Automated Traffic Analysis," in *ACM SIGSAC Conference on Computer and Communications Security (CCS)*, 2021, pp. 3366–3383.
- [74] T. Shapira and Y. Shavitt, "Flowpic: A generic representation for encrypted traffic classification and applications identification," *IEEE Transactions on Network and Service Management*, vol. 18, no. 2, pp. 1218–1232, 2021.
- [75] N. Sivaroopan, D. Bandara, C. Madarasingha, G. Jourjon, A. P. Jayasumana, and K. Thilakarathna, "NetDiffus: Network Traffic Generation by Diffusion Models through Time-Series Imaging," *Computer Networks*, vol. 251, p. 110616, 2024.
- [76] R. F. Bikmukhamedov and A. F. Nadeev, "Multi-Class Network Traffic Generators and Classifiers Based on Neural Networks," in *Systems of Signals Generating and Processing in the Field of on Board Communications*, 2021, pp. 1–7.
- [77] D. K. Kholgh and P. Kostakos, "PAC-GPT: A Novel Approach to Generating Synthetic Network Traffic With GPT-3," *IEEE Access*, vol. 11, pp. 114 936–114 951, 2023.
- [78] Q. Wang, C. Qian, X. Li, Z. Yao, and H. Shao, "LENS: A Foundation Model for Network Traffic," *arXiv preprint arXiv:2402.03646*, 2024.
- [79] J. Qu, X. Ma, and J. Li, "TrafficGPT: Breaking the Token Barrier for Efficient Long Traffic Analysis and Generation," *arXiv preprint arXiv:2403.05822*, 2024.
- [80] A. Chu, X. Jiang, S. Liu, A. Bhagoji, F. Bronzino, P. Schmitt, and N. Feamster, "Feasibility of state space models for network traffic generation," *arXiv preprint arXiv:2406.02784*, 2024.
- [81] S. Zhang, T. Li, D. Jin, and Y. Li, "NetDiff: A Service-Guided Hierarchical Diffusion Model for Network Flow Trace Generation," *Proc. ACM Netw.*, vol. 2, no. CoNEXT3, Aug. 2024.
- [82] F. Li, H. Wu, and J. Zhang, "Lightweight diffusion model for synthesizing malicious network traffic," in *NAECON 2024 - IEEE National Aerospace and Electronics Conference*, 2024, pp. 409–413.
- [83] M. Wolf, J. Tritscher, D. Landes, A. Hotho, and D. Schlör, "Benchmarking of Synthetic Network Data: Reviewing Challenges and Approaches," *Computers & Security*, vol. 145, p. 103993, 2024.
- [84] Z. Wang and T. Oates, "Imaging Time-series to Improve Classification and Imputation," in *24th International Conference on Artificial Intelligence (IJCAI)*, *ML Track*, 2015, pp. 3939–3945.
- [85] S. Guthula, N. Battula, R. Beltiukov, W. Guo, and A. Gupta, "net-

- Found: Foundation Model for Network Security,” *arXiv preprint arXiv:2310.17025*, 2023.
- [86] A. Sarabi, T. Yin, and M. Liu, “An LLM-based Framework for Fingerprinting Internet-connected Devices,” in *ACM on Internet Measurement Conference (IMC)*, 2023, p. 478–484.
- [87] T. Liu, X. Ma, L. Liu, X. Liu, Y. Zhao, N. Hu, and K. Z. Ghafoor, “LAMBERT: Leveraging Attention Mechanisms to Improve the BERT Fine-Tuning Model for Encrypted Traffic Classification,” *Mathematics*, vol. 12, no. 11, 2024.
- [88] H. Li, Y. Zhang, M. Gu, J. Bai, Z. Xiao, and Y. Wang, “Encrypted Traffic Classification Framework Based on Albert,” in *IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium*, 2024, pp. 10019–10024.
- [89] M. Nam, S. Park, and D. S. Kim, “Intrusion Detection Method using Bi-directional GPT for In-vehicle Controller Area Networks,” *IEEE Access*, vol. 9, pp. 124 931–124 944, 2021.
- [90] K. Yu, L. Tan, S. Mumtaz, S. Al-Rubaye, A. Al-Dulaimi, A. K. Bashir, and F. A. Khan, “Securing Critical Infrastructures: Deep-Learning-Based Threat Detection in IIoT,” *IEEE Communications Magazine*, vol. 59, no. 10, pp. 76–82, 2021.
- [91] Y. Li, X. Yuan, and W. Li, “An Extreme Semi-supervised Framework Based on Transformer for Network Intrusion Detection,” in *31st ACM International Conference on Information & Knowledge Management (CIKM)*, 2022, pp. 4204–4208.
- [92] Y. E. Seyyar, A. G. Yavuz, and H. M. Ünver, “An Attack Detection Framework Based on BERT and Deep Learning,” *IEEE Access*, vol. 10, pp. 68 633–68 644, 2022.
- [93] C. M. K. Ho, K.-C. Yow, Z. Zhu, and S. Aravamuthan, “Network Intrusion Detection via Flow-to-Image Conversion and Vision Transformer Classification,” *IEEE Access*, vol. 10, pp. 97 780–97 793, 2023.
- [94] Z. Wu, H. Zhang, P. Wang, and Z. Sun, “RTIDS: A Robust Transformer-Based Approach for Intrusion Detection System,” *IEEE Access*, vol. 10, pp. 64 375–64 387, 2022.
- [95] A. Ghourabi, “A Security Model Based on LightGBM and Transformer to Protect Healthcare Systems from Cyberattacks,” *IEEE Access*, vol. 10, pp. 48 890–48 903, 2022.
- [96] H. Lai, “Intrusion Detection Technology Based on Large Language Models,” in *IEEE International Conference on Evolutionary Algorithms and Soft Computing Techniques (EASCT)*, 2023, pp. 1–5.
- [97] T. Ali and P. Kostakos, “HuntGPT: Integrating Machine Learning-Based Anomaly Detection and Explainable AI with Large Language Models (LLMs),” *arXiv e-prints*, p. arXiv:2309.16021, 2023.
- [98] S. Ullah, J. Ahmad, M. A. Khan, M. S. Alshehri, W. Boulila, A. Koubaa, S. U. Jan, and M. M. I. Ch, “TNN-IDS: Transformer Neural Network-based Intrusion Detection System for MQTT-enabled IIoT Networks,” *Computer Networks*, vol. 237, p. 110072, 2023.
- [99] W. Wang, S. Jian, Y. Tan, Q. Wu, and C. Huang, “Robust Unsupervised Network Intrusion Detection with Self-supervised Masked Context Reconstruction,” *Computers & Security*, vol. 128, p. 103131, 2023.
- [100] Z. Wang, J. Li, S. Yang, X. Luo, D. Li, and S. Mahmoodi, “A Lightweight IIoT Intrusion Detection Model based on Improved BERT-of-Theseus,” *Expert Systems with Applications*, vol. 238, p. 122045, 2024.
- [101] F. S. Melfías, T. F. Ribeiro, C. Rabadão, L. Santos, and R. L. d. C. Costa, “GPT and Interpolation-based Data Augmentation for Multiclass Intrusion Detection in IIoT,” *IEEE Access*, 2024.
- [102] F. Setianto, E. Tsani, F. Sadiq, G. Domalis, D. Tsakalidis, and P. Kostakos, “GPT-2C: A Parser for HoneyPot Logs using Large Pre-trained Language Models,” in *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 2021, pp. 649–653.
- [103] H. Ott, J. Bogatinovski, A. Acker, S. Nedelkoski, and O. Kao, “Robust and Transferable Anomaly Detection in Log Data using Pre-Trained Language Models,” in *IEEE/ACM International Workshop on Cloud Intelligence (CloudIntelligence)*, 2021, pp. 19–24.
- [104] J. Pan, S. L. Wong, and Y. Yuan, “RAGLog: Log Anomaly Detection using Retrieval Augmented Generation,” *arXiv preprint arXiv:2311.05261*, 2023.
- [105] J. Qi, S. Huang, Z. Luan, S. Yang, C. Fung, H. Yang, D. Qian, J. Shang, Z. Xiao, and Z. Wu, “LogGPT: Exploring ChatGPT for Log-based Anomaly Detection,” in *IEEE International Conference on High Performance Computing & Communications, Data Science & Systems, Smart City & Dependability in Sensor, Cloud & Big Data Systems & Application (HPCC/DSS/SmartCity/DependSys)*, 2023, pp. 273–280.
- [106] Z. Jiang, J. Liu, Z. Chen, Y. Li, J. Huang, Y. Huo, P. He, J. Ge, and M. R. Lyu, “LILAC: Log Parsing using LLMs with Adaptive Parsing Cache,” *Proceedings of the ACM on Software Engineering*, vol. 1, pp. 137–160, 2024.
- [107] Y. Ji, J. Han, Y. Zhao, S. Zhang, and Z. Gong, “Log Anomaly Detection Through GPT-2 for Large Scale Systems,” *ZTE Communications*, vol. 21, no. 3, p. 70, 2023.
- [108] P. Mudgal and R. Wouhaybi, “An Assessment of ChatGPT on Log Data,” in *1st International Conference on AI-generated Content (AIGC)*, 2023, pp. 148–169.
- [109] Y. Sun, Y. Chen, H. Zhao, and S. Peng, “Design and Development of a Log Management System Based on Cloud Native Architecture,” in *9th IEEE International Conference on Systems and Informatics (ICSAI)*, 2023, pp. 1–6.
- [110] T. Vörös, S. P. Bergeron, and K. Berlin, “Web Content Filtering through Knowledge Distillation of Large Language Models,” in *IEEE International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, 2023, pp. 357–361.
- [111] P. Balasubramanian, J. Seby, and P. Kostakos, “CYGENT: A Cyber-security Conversational Agent with Log Summarization Powered by GPT-3,” *arXiv preprint arXiv:2403.17160*, 2024.
- [112] E. Karlsen, X. Luo, N. Zincir-Heywood, and M. Heywood, “Large Language Models and Unsupervised Feature Learning: Implications for Log Analysis,” *Annals of Telecommunications*, pp. 1–19, 2024.
- [113] B. Meyuhass, A. Bremner-Barr, and T. Shapira, “IoT device labeling using large language models,” *arXiv preprint arXiv:2403.01586*, 2024.
- [114] Y. Tian and Z. Li, “Dom-BERT: Detecting Malicious Domains with Pre-training Model,” in *International Conference on Passive and Active Network Measurement (PAM)*. Springer, 2024, pp. 133–158.
- [115] C. Almodovar, F. Sabrina, S. Karimi, and S. Azad, “Logfit: Log anomaly detection using fine-tuned language models,” *IEEE Transactions on Network and Service Management*, 2024.
- [116] S. Soman and R. HG, “Observations on LLMs for Telecom Domain: Capabilities and Limitations,” in *3rd ACM International Conference on AI-ML Systems (AIMLSys)*, 2023, pp. 1–5.
- [117] C. Wang, M. Scazzariello, A. Farshin, D. Kostic, and M. Chiesa, “Making network configuration human friendly,” *arXiv preprint arXiv:2309.06342*, 2023.
- [118] S. K. Mani, Y. Zhou, K. Hsieh, S. Segarra, T. Eberl, E. Azulai, I. Frizler, R. Chandra, and S. Kandula, “Enhancing Network Management Using Code Generated by Large Language Models,” in *22nd ACM Workshop on Hot Topics in Networks (HotNets)*, 2023, pp. 196–204.
- [119] R. Mondal, A. Tang, R. Beckett, T. Millstein, and G. Varghese, “What Do LLMs Need to Synthesize Correct Router Configurations?” in *22nd ACM Workshop on Hot Topics in Networks (HotNets)*, 2023, pp. 189–195.
- [120] S. Roychowdhury, N. Jain, and S. Soman, “Unlocking Telecom Domain Knowledge Using LLMs,” in *16th IEEE International Conference on Communication Systems & NETWORKS (COMSNETS)*, 2024, pp. 267–269.
- [121] Y. Shen, J. Shao, X. Zhang, Z. Lin, H. Pan, D. Li, J. Zhang, and K. B. Letaief, “Large Language Models Empowered Autonomous Edge AI for Connected Intelligence,” *IEEE Communications Magazine*, 2024.
- [122] M. Duclos, I. A. Fernandez, K. Moore, S. Mittal, and E. Ziegler, “Utilizing Large Language Models to Translate RFC Protocol Specifications to CPISA Definitions,” *arXiv preprint arXiv:2402.00890*, 2024.
- [123] T. Ahmed, N. Piovesan, A. De Domenico, and S. Choudhury, “Linguistic Intelligence in Large Language Models for Telecommunications,” *arXiv preprint arXiv:2402.15818*, 2024.
- [124] N. Piovesan, A. De Domenico, and F. Ayed, “Telecom Language Models: Must They Be Large?” *arXiv preprint arXiv:2403.04666*, 2024.
- [125] H. Ghasemirahni, A. Farshin, M. Scazzariello, M. Chiesa, and D. Kostić, “Deploying Stateful Network Functions Efficiently using Large Language Models,” in *4th ACM-EUROSYS Workshop on Machine Learning and Systems (EuroMLSys)*, 2024, pp. 28–38.
- [126] O. Erak, N. Alabbasi, O. Alhussein, I. Lotfi, A. Hussein, S. Muhaidat, and M. Debbah, “Leveraging Fine-Tuned Retrieval-Augmented Generation with Long-Context Support: For 3GPP Standards,” *arXiv preprint arXiv:2408.11775*, 2024.
- [127] F. Ayed, A. Maatouk, N. Piovesan, A. De Domenico, M. Debbah, and Z.-Q. Luo, “Hermes: A large language model framework on the journey to autonomous networks,” *arXiv preprint arXiv:2411.06490*, 2024.
- [128] M. Ahmed, S. Byreddy, A. Nutakki, L. F. Sikos, and P. Haskell-Dowland, “ECU-IoHT: A Dataset for Analyzing Cyberattacks in Internet of Health Things,” *Ad Hoc Networks*, vol. 122, p. 102621, 2021.
- [129] H. Jin, X. Han, J. Yang, Z. Jiang, Z. Liu, C.-Y. Chang, H. Chen, and X. Hu, “LLM Maybe LongLM: Self-Extend LLM Context Window Without Tuning,” *arXiv preprint arXiv:2401.01325*, 2024.
- [130] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. R. Salakhutdinov, and Q. V.

- Le, "XLNet: Generalized Autoregressive Pretraining for Language Understanding," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 32, 2019.
- [131] Y. Yao, J. Duan, K. Xu, Y. Cai, Z. Sun, and Y. Zhang, "A Survey on Large Language Model (LLM) Security and Privacy: The Good, the Bad, and the Ugly," *High-Confidence Computing*, p. 100211, 2024.
- [132] M. Tavallaei, E. Bagheri, W. Lu, and A. A. Ghorbani, "A Detailed Analysis of the KDD CUP 99 Data Set," in *IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA)*, 2009, pp. 1–6.
- [133] C. Raissi, J. Brissaud, G. Dray, P. Poncelet, M. Roche, and M. Teisseire, "Web Analyzing Traffic Challenge: Description and Results," in *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD)*, 2007, pp. 47–52.
- [134] "CSIC 2010 Web Application Attacks." [Online]. Available: <https://www.kaggle.com/datasets/ispangler/csic-2010-web-application-attacks>
- [135] "HttpParamsDataset." [Online]. Available: <https://www.kaggle.com/datasets/evg3n1j/httpparamsdataset>
- [136] N. Moustafa and J. Slay, "UNSW-NB15: A Comprehensive Data Set for Network Intrusion Detection Systems (UNSW-NB15 Network Data Set)," in *IEEE Military Communications and Information Systems Conference (MilCIS)*, 2015, pp. 1–6.
- [137] M. Miettinen, S. Marchal, I. Hafeez, N. Asokan, A.-R. Sadeghi, and S. Tarkoma, "IoT SENTINEL: Automated Device-Type Identification for Security Enforcement in IoT," in *IEEE 37th International Conference on Distributed Computing Systems (ICDCS)*, 2017, pp. 2177–2184.
- [138] G. Draper-Gil, A. H. Lashkari, M. S. I. Mamun, and A. A. Ghorbani, "Characterization of Encrypted and VPN Traffic using Time-related Features," in *International Conference on Information Systems Security and Privacy (ICISSP)*, 2016, pp. 407–414.
- [139] A. H. Lashkari, G. D. Gil, M. S. I. Mamun, and A. A. Ghorbani, "Characterization of Tor Traffic using Time based Features," in *International Conference on Information Systems Security and Privacy (ICISSP)*, vol. 2, 2017, pp. 253–262.
- [140] A. Sivanathan, H. H. Gharakheili, F. Loi, A. Radford, C. Wijenayake, A. Vishwanath, and V. Sivaraman, "Classifying IoT Devices in Smart Environments Using Network Traffic Characteristics," *IEEE Transactions on Mobile Computing*, vol. 18, no. 8, pp. 1745–1759, 2018.
- [141] "USTC-TFC2016." [Online]. Available: <https://www.kaggle.com/datasets/randasrou/ustctfc2016>
- [142] I. Sharafaldin, A. H. Lashkari, A. A. Ghorbani *et al.*, "Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization," *ICISSp*, vol. 1, pp. 108–116, 2018.
- [143] P. Sirinam, M. Imani, M. Juarez, and M. Wright, "Deep Fingerprinting: Undermining Website Fingerprinting Defenses with Deep Learning," in *ACM SIGSAC Conference on Computer and Communications Security (CCS)*, 2018, pp. 1928–1943.
- [144] "LabeledFlows2017 Dataset," <https://www.kaggle.com/datasets/jsrojas/ip-network-traffic-flows-labeled-with-87-apps>.
- [145] T. Van Ede, R. Bortolameotti, A. Continella, J. Ren, D. J. Dubois, M. Lindorfer, D. Choffnes, M. Van Steen, and A. Peter, "FlowPrint: Semi-Supervised Mobile-App Fingerprinting on Encrypted Network Traffic," in *Network and Distributed System Security symposium (NDSS)*, vol. 27, 2020.
- [146] X. V. Lin, C. Wang, L. Zettlemoyer, and M. D. Ernst, "NL2Bash: A Corpus and Semantic Parser for Natural Language Interface to the Linux Operating System," in *Language Resource and Evaluation Conference (LREC)*, 2018.
- [147] I. Sharafaldin, A. H. Lashkari, S. Hakak, and A. A. Ghorbani, "Developing Realistic Distributed Denial of Service (DDoS) Attack Dataset and Taxonomy," in *IEEE International Carnahan Conference on Security Technology (ICCST)*, 2019, pp. 1–8.
- [148] R. Perdisci, T. Papastergiou, O. Alrawi, and M. Antonakakis, "IoTFinder: Efficient Large-Scale Identification of IoT Devices via Passive DNS Traffic Analysis," in *IEEE European Symposium on Security and Privacy (EuroS&P)*, 2020, pp. 474–489.
- [149] F. Zaker, "Online Shopping Store - Web Server Logs," 2019. [Online]. Available: <https://www.kaggle.com/datasets/eliassabbas/web-server-access-logs>
- [150] "LabeledFlows2019 Dataset," <https://www.kaggle.com/datasets/jsrojas/labeled-network-traffic-flows-114-applications>.
- [151] M. A. A. Hilmi, K. A. Cahyanto, and M. Mustamiin, "Apache Web Server - Access Log Pre-processing for Web Intrusion Detection," 2020. [Online]. Available: <https://dx.doi.org/10.25126/jtiik.2022924107>
- [152] M. MontazeriShatoori, L. Davidson, G. Kaur, and A. H. Lashkari, "Detection of DoH Tunnels using Time-series Classification of Encrypted Traffic," in *IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech)*, 2020, pp. 63–70.
- [153] U. Sedlar, M. Kren, L. Štefanič Južnič, and M. Volk, "CyberLab HoneyNet Dataset," 2020. [Online]. Available: <https://doi.org/10.5281/zenodo.3687527>
- [154] J. Zhu, S. He, P. He, J. Liu, and M. R. Lyu, "Loghub: A Large Collection of System Log Datasets for AI-driven Log Analytics," in *IEEE 34th International Symposium on Software Reliability Engineering (ISSRE)*, 2023, pp. 355–366.
- [155] A. Brenner-Barr, B. Meyuhass, and R. Shister, "One MUD to Rule Them All: IoT Location Impact," in *IEEE/IFIP Network Operations and Management Symposium (NOMS)*, 2022, pp. 1–5.
- [156] H. Hindy, E. Bayne, M. Bures, R. Atkinson, C. Tachtatzis, and X. Bellekens, "Machine Learning based IoT Intrusion Detection System: An MQTT Case Study (MQTT-IoT-IDS2020 Dataset)," in *International Networking Conference*. Springer, 2020, pp. 73–84.
- [157] N. Moustafa, "A New Distributed Architecture for Evaluating AI-based Security Systems at the Edge: Network TON\_IoT Datasets," *Sustainable Cities and Society*, vol. 72, p. 102994, 2021.
- [158] "fwaf-dataset." [Online]. Available: <https://www.kaggle.com/datasets/evg3n1j/fwaf-dataset>
- [159] E. S. Consortium, "Dataset of EU SPIRIT Project," 2021. [Online]. Available: <https://doi.org/10.5281/zenodo.4767861>
- [160] S. Dadkhah, H. Mahdikhani, P. K. Danso, A. Zohourian, K. A. Truong, and A. A. Ghorbani, "Towards the Development of a Realistic Multidimensional IoT Profiling Dataset," in *19th Annual International Conference on Privacy, Security & Trust (PST)*, 2022, pp. 1–11.
- [161] M. A. Ferrag, O. Friha, D. Hamouda, L. Maglaras, and H. Janicke, "Edge-IIoTset: A New Comprehensive Realistic Cyber Security Dataset of IoT and IIoT Applications for Centralized and Federated Learning," *IEEE Access*, vol. 10, pp. 40281–40306, 2022.
- [162] E. C. P. Neto, S. Dadkhah, R. Ferreira, A. Zohourian, R. Lu, and A. A. Ghorbani, "CICIoT2023: A Real-Time Dataset and Benchmark for Large-Scale Attacks in IoT Environment," *Sensors*, vol. 23, no. 13, p. 5941, 2023.
- [163] I. Karim, K. S. Mubasshir, M. M. Rahman, and E. Bertino, "SPEC5G: A Dataset for 5G Cellular Network Protocol Analysis," in *13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics (IJCNLP-AACL)*, 2023.
- [164] A. Maatouk, F. Ayed, N. Piovesan, A. De Domenico, M. Debbah, and Z.-Q. Luo, "TeleQnA: A Benchmark Dataset to Assess Large Language Models Telecommunications Knowledge," *arXiv preprint arXiv:2310.15051*, 2023.
- [165] Z. Durumeric, D. Adrian, A. Mirian, M. Bailey, and J. A. Halderman, "A Search Engine Backed by Internet-Wide Scanning," in *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security (CCS)*, 2015, pp. 542–553.
- [166] S. Zhu, J. Lu, B. Lyu, T. Pan, C. Jia, X. Cheng, D. Kang, Y. Lv, F. Yang, X. Xue *et al.*, "Zoonet: A Proactive Telemetry System for Large-Scale Cloud Networks," in *Proceedings of the 18th International Conference on emerging Networking EXperiments and Technologies*, 2022, pp. 321–336.
- [167] M. Yu, "Network Telemetry: Towards a Top-Down Approach," *ACM SIGCOMM Computer Communication Review*, vol. 49, no. 1, pp. 11–17, 2019.
- [168] A. D'Alconzo, I. Drago, A. Morichetta, M. Mellia, and P. Casas, "A Survey on Big Data for Network Traffic Monitoring and Analysis," *IEEE Transactions on Network and Service Management*, vol. 16, no. 3, pp. 800–813, 2019.
- [169] C. Qian, X. Li, Q. Wang, G. Zhou, and H. Shao, "NetBench: A Large-Scale and Comprehensive Network Traffic Benchmark Dataset for Foundation Models," *arXiv preprint arXiv:2403.10319*, 2024.
- [170] A. S. Jacobs, R. Beltiukov, W. Willinger, R. A. Ferreira, A. Gupta, and L. Z. Granville, "AI/ML for network security: The emperor has no clothes," in *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security*, 2022, pp. 1537–1551.
- [171] G. Engelen, V. Rimmer, and W. Joosen, "Troubleshooting an Intrusion Detection Dataset: the CICIDS2017 Case Study," in *IEEE Security and Privacy Workshops (SPW)*, 2021, pp. 7–12.
- [172] C. Zhang, Y. Xie, H. Bai, B. Yu, W. Li, and Y. Gao, "A Survey on Federated Learning," *Knowledge-Based Systems*, vol. 216, p. 106775, 2021.
- [173] L. Li, Y. Fan, M. Tse, and K.-Y. Lin, "A Review of Applications in

Federated Learning,” *Computers & Industrial Engineering*, vol. 149, p. 106854, 2020.

- [174] X. Huang, P. Li, H. Du, J. Kang, D. Niyato, D. I. Kim, and Y. Wu, “Federated Learning-Empowered AI-Generated Content in Wireless Networks,” *IEEE Network*, 2024.
- [175] R. Schwartz, J. Dodge, N. A. Smith, and O. Etzioni, “Green AI,” *Communications of the ACM*, vol. 63, no. 12, pp. 54–63, 2020.
- [176] P. P. Ray, “A Review on TinyML: State-of-the-art and Prospects,” *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 4, pp. 1595–1623, 2022.
- [177] L. Dutta and S. Bharali, “TinyML Meets IoT: A Comprehensive Survey,” *Internet of Things*, vol. 16, p. 100461, 2021.
- [178] C. Ciliberto, M. Herberster, A. D. Ialongo, M. Pontil, A. Rocchetto, S. Severini, and L. Wossnig, “Quantum Machine Learning: A Classical Perspective,” *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 474, no. 2209, p. 20170551, 2018.
- [179] V. Dunjko and H. J. Briegel, “Machine Learning & Artificial Intelligence in the Quantum Domain: A Review of Recent Progress,” *Reports on Progress in Physics*, vol. 81, no. 7, p. 074001, 2018.
- [180] H. Du, R. Zhang, Y. Liu, J. Wang, Y. Lin, Z. Li, D. Niyato, J. Kang, Z. Xiong, S. Cui *et al.*, “Beyond deep reinforcement learning: A tutorial on generative diffusion models in network optimization,” *arXiv preprint arXiv:2308.05384*, 2023.
- [181] —, “Enhancing deep reinforcement learning: A tutorial on generative diffusion models in network optimization,” *IEEE Communications Surveys & Tutorials*, 2024.
- [182] F. Xu, H. Uszkoreit, Y. Du, W. Fan, D. Zhao, and J. Zhu, “Explainable AI: A Brief Survey on History, Research Areas, Approaches and Challenges,” in *Natural language processing and Chinese computing: 8th cCF international conference*. Springer, 2019, pp. 563–574.
- [183] R. Dwivedi, D. Dave, H. Naik, S. Singhal, R. Omer, P. Patel, B. Qian, Z. Wen, T. Shah, G. Morgan *et al.*, “Explainable AI (XAI): Core Ideas, Techniques, and Solutions,” *ACM Computing Surveys*, vol. 55, no. 9, pp. 1–33, 2023.
- [184] A. Nascita, G. Aceto, D. Ciuonzo, A. Montieri, V. Persico, and A. Pescapé, “A Survey on Explainable Artificial Intelligence for Internet Traffic Classification and Prediction, and Intrusion Detection,” *IEEE Communications Surveys & Tutorials*, 2024.
- [185] J. Kaddour, A. Lynch, Q. Liu, M. J. Kusner, and R. Silva, “Causal Machine Learning: A Survey and Open Problems,” *arXiv preprint arXiv:2206.15475*, 2022.
- [186] M. K. Sarker, L. Zhou, A. Eberhart, and P. Hitzler, “Neuro-Symbolic Artificial Intelligence,” *AI Communications*, vol. 34, no. 3, pp. 197–209, 2021.
- [187] P. Hitzler and M. K. Sarker, *Neuro-Symbolic Artificial Intelligence: The State of the Art*. IOS press, 2022.
- [188] Z. Zheng, S. Xie, H.-N. Dai, X. Chen, and H. Wang, “Blockchain Challenges and Opportunities: A Survey,” *International Journal of Web and Grid Services*, vol. 14, no. 4, pp. 352–375, 2018.
- [189] M. Pilkington, “Blockchain Technology: Principles and Applications,” in *Research handbook on digital transformations*. Edward Elgar Publishing, 2016, pp. 225–253.



**Giampaolo Bovenzi** is an Assistant Professor at DIETI of the University of Napoli Federico II, since October 2023. He received his Ph.D. degree at the same University in June 2022. His research interests focus on (anonymized and encrypted) traffic classification, network security (with a focus on IoT), and blockchain. He has co-authored more than 20 papers in international journals and conference proceedings.



**Francesco Cerasuolo** is a Ph.D. student at DIETI of the University of Napoli Federico II. He received his M.S. Laurea Degree in Computer Engineering in July 2022 from the same University. His research interests include traffic classification, machine and deep learning, and class incremental learning.



**Domenico Ciuonzo (S'11-M'14-SM'16)** is a Tenure-Track Professor at the University of Napoli Federico II. He holds a Ph.D. from the University of Campania Luigi Vanvitelli. He is the recipient of two Best Paper awards (IEEE ICCCS 2019 and Elsevier ComNet 2020), the 2019 IEEE AESS Exceptional Service award, the 2020 IEEE SENSORS COUNCIL Early-Career Technical Achievement award and the 2021 IEEE AESS Early-Career Award. His research interests include data fusion, network analytics, IoT, and AI.



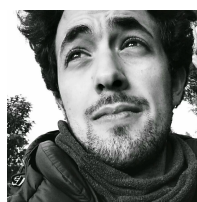
**Davide Di Monda** is a Ph.D. student at IMT School for Advanced Studies Lucca and University of Napoli Federico II since December 2022. He received his M.S. Laurea Degree (summa cum laude) in Computer Engineering in July 2022 from the University of Napoli Federico II. His research interests include cybersecurity, attack classification, and blockchain.



**Idio Guarino** is a postdoctoral researcher in the Department of Computer Science of the University of Verona since April 2024. He received his Ph.D. in Information Technology and Electrical Engineering in February 2024 and his M.S. degree in Computer Engineering in July 2020, both from the University of Napoli Federico II. His research focuses on analyzing network traffic via AI to develop innovative methodologies for network traffic classification and prediction.



**Antonio Montieri** is an Assistant Professor at the University of Napoli Federico II, where he earned his Ph.D. in Information Technology and Electrical Engineering in April 2020. His research focuses on network security, traffic classification, modeling and prediction, and explainable AI and generative AI for networking and traffic analysis. He has co-authored over 50 papers in leading international venues and received various awards, including the Computer Networks 2020 Best Paper Award.



**Valerio Persico** is an Associate Professor at the University of Napoli Federico II, where he received the Ph.D. in Computer and Automation Engineering in 2016. His work concerns network measurements, traffic analysis, cloud-network monitoring, and Internet path tracing. He has co-authored more than 70 papers within international journals and conference proceedings and is the recipient of several awards, including IEEE ISCC 2022, IEEE ICCCS 2019, and IEEE CSIM 2018 Best Paper awards.



**Antonio Pescapé (SM'09)** is a Full Professor of Computer Engineering at the University of Napoli Federico II. His work focuses on Internet technologies, more precisely on measurement, monitoring, and analysis of the Internet, and on AI for networking. He has co-authored more than 200 conference and journal papers and he is the recipient of several research awards. Also, he has served as an independent reviewer/evaluator of research projects/project proposals co-funded by a number of governments and agencies.