MetaWatch: Trends, Challenges, and Future of Network Intrusion Detection in the Metaverse

Ebuka Chinaechetam Nkoro, Judith Nkechinyere Njoku, *Member, IEEE*, Cosmas Ifeanyi Nwakanma, *Senior Member, IEEE*, Jae Min Lee, *Member, IEEE*, and Dong-Seong Kim, *Senior Member, IEEE*

***Abstract*—As the Metaverse progresses, its security measures must evolve to safeguard users from cyberattacks. To this end, Artificial Intelligence (AI)—powered Network Intrusion Detection Systems (NIDSs) have been implemented at the network layer to detect and respond to threats in real-time. Our survey—*MetaWatch* provides a comprehensive overview and analysis of relevant literature, focusing on studies that have explored NIDSs in the Metaverse. Unlike other surveys that have addressed Metaverse security issues more broadly, *MetaWatch* specifically provides a taxonomy of detection paradigms and defense mechanisms within Metaverse NIDS from 2021 to 2024. Additionally, *MetaWatch* identifies and discusses the key challenges and problems that impact the development of trustworthy, explainable, scalable, and robust Metaverse NIDS. Overall, the survey provides readers with a concise and informative summary of NIDS issues in the Metaverse and highlights open security problems worth exploring.**

***Index Terms*—5G/6G, artificial intelligence (AI), cybersecurity, explainable AI (XAI), Internet of Things (IoT), metaverse, network intrusion detection system (NIDS).**

I. Introduction

THE TERM *Metaverse* [1] refers to a simulated 3-D virtual world, including Virtual Reality (VR), Mixed Reality (MR), and Extended Reality (XR), all connected to physical (actual) objects. Following the huge investments and diverse applications of metaverse technologies [2], there have been several issues of unauthorized breaches of confidentiality, integrity, and availability (CIA) of Metaverse networks, which limits access to its resources and data [3], [4]. Stealthy Metaverse cyberattacks like man-in-the-room attacks [5], phishing attacks, gesture recognition, and social engineering attacks have been reported by users and various security researchers [6].

A recent example of an Australian Metaverse platform—STEPN, witnessed multiple Distributed Denial of Service (DDoS) attacks with over 25 million messages proliferated by unidentified bad actors, which led to hours of server shutdowns, poor user experience, and eventually affected the company's reputation, and productivity [7]. These attacks pose diverse risks of user privacy, asset loss, organizational reputation, and loss of interest in the Metaverse.

The scope of this review is narrowed to Network Intrusion Detection Systems (NIDS) within the Metaverse, as opposed to a broader field of intrusion detection involving host intrusion detection deployment methods. The focus on Metaverse NIDS is justified by the critical role played by high-volume heterogeneous *network traffic*, particularly within Fifth/Sixth Generation (5G/6G) networks, which makes the immersive Metaverse experience vulnerable to sophisticated attacks and vulnerabilities. Moreover, intrusion detection at the network layer is chosen as a subject focus due to its dynamic and intelligent capabilities for monitoring and detecting malicious activities within the network compared to host intrusion methods, such as system log analysis, file access attempts, and server process monitoring, which are typically resource-intensive, static, and still suffer delayed threat detection [8], [9].

A. Background

The attack surface of the Metaverse is predicated on the vulnerabilities of the Internet of Things (IoT)-enabled haptic devices that collect body, gesture, gait, location, and personal multisensory information that bad actors can leverage to carry out malicious attacks [5], [10]. Other open vulnerabilities can emanate from careless users or advanced persistent groups who aim to manipulate these devices to steal users' personal information, distort optimal network communications, and request ransoms for successful exploits.

A potential solution toward mitigating such network-based attacks in Metaverse is using artificial intelligence (AI)—enabled NIDSs [11]. NIDSs utilize machine learning (ML) algorithms to monitor and emulate the eye of a skilled security analyst to detect and respond intelligently and successfully identify threats. The NIDS triggers security policies

Received 13 March 2024; revised 6 September 2024, 17 March 2025, and 16 April 2025; accepted 6 May 2025. Date of publication 9 May 2025; date of current version 8 August 2025. This work was supported in part by the Innovative Human Resource Development for Local Intellectualization Program through the IITP Grant funded by the Korea Government (MSIT) under Grant IITP-2025-RS-2020-II201612 (25%); in part by the Priority Research Centers Program through the NRF funded by the MEST under Grant 2018R1A6A1A03024003 (25%); in part by MSIT, South Korea, through the ITRC Support Program under Grant IITP-2025-RS-2024-00438430 (25%); and in part by the Basic Science Research Program through the NRF funded by the MOE under Grant 2022R1I1A3071844 (25%). *(Corresponding author: Dong-Seong Kim.)*

Ebuka Chinaechetam Nkoro is with the ICT Convergence Research Center, Kumoh National Institute of Technology, Gumi 39177, South Korea.

Judith Nkechinyere Njoku, Jae Min Lee, and Dong-Seong Kim are with the Department of IT Convergence Engineering, Kumoh National Institute of Technology, Gumi 39177, South Korea (e-mail: dskim@kumoh.ac.kr).

Cosmas Ifeanyi Nwakanma is with the Lane Department of Computer Science and Electrical Engineering, West Virginia University, Morgantown, WV 26506 USA.

Digital Object Identifier 10.1109/JIOT.2025.3568477

to defend against real-time and future threats in the Metaverse network.

Despite the auspicious efforts by Metaverse NIDS researchers [6], [11], [12], [13], there is a limited summary of the trends, essential methods, and limitations of general approaches employed for Metaverse NIDSs. Current Metaverse NIDS challenges and trends worth reviewing include dataset bottlenecks, modeling methods, adversarial defense techniques, privacy preservation demands, and the need for eXplainable AI (XAI) methods that security researchers can utilize to improve Metaverse NIDS postures.

B. Motivation

The rapid evolution of the Metaverse, encompassing diverse applications, haptic devices, and virtual environments, introduces significant security and privacy challenges. Stakeholders in Metaverse cybersecurity, including Metaverse developers, security researchers, and end-users, are increasingly concerned about the potential vulnerabilities that could undermine the trust and safety of these virtual spaces. Existing literature, such as the surveys by Wang et al. [4], Huang et al. [14], Di Pietro and Cresci [15], Chen et al. [16], and Jaber [17], has extensively reviewed the broad spectrum of security and privacy challenges within the Metaverse. However, a critical gap remains in the systematic analysis of AI-enabled Metaverse NIDS methods, a rapidly emerging area of interest for both developers and researchers aiming to enhance the security infrastructure of the Metaverse.

This study addresses this gap by comprehensively evaluating and reviewing AI-enabled Metaverse NIDS approaches from 2021 to 2024. We offer a detailed taxonomy highlighting the strengths and limitations of current intrusion detection techniques and suggest future research directions to build a more trustworthy Metaverse NIDS. This research will significantly benefit developers in designing more secure Metaverse platforms, aid researchers in focusing their efforts on the most pressing security challenges, and ultimately contribute to a safer and more reliable user experience in the Metaverse. Our work is motivated by the need to provide stakeholders with a systematic and practical analysis of the current state of AI-enabled NIDS in the Metaverse, ensuring that a thorough understanding of existing methods and challenges informs future developments. A list of abbreviations is provided in Table I to improve readability.

TABLE I
ABBREVIATIONS USED IN THIS SURVEY

Abbreviations	Meaning
AE	Auto Encoder
AI	Artificial Intelligence
CIA	Confidentiality, Integrity and Availability
CNN	Convolutional Neural Network
CDA	Community Detection Algorithm
CPS	Cyber-Physical System
DT	Decision Tree
DT	Digital Twins
DL	Deep Learning
DoS	Denial of Service
DDoS	Distributed Denial of Service attacks
e-SIM	Electronic Subscriber Identification Module
ERC	Ethereum Request for Comment
FL	Federated Learning
GAN	Generative Adversarial Networks
GUI	Graphic User Interface
HMD	Head Mounted Display
HIDS	Host-based Intrusion Detection System
I/O	Input Output
IoT	Internet of Things
KPCA	Kernel Principal Component Analysis
LLM	Large Language Model
LIME	Local Interpretable Model-Agnostic Explanations
LR	Logistic Regression
ML	Machine Learning
MR	Mixed Reality
MTD	Moving Target Defence
NIDS	Network Intrusion Detection System
PCA	Principal Component Analysis
PoA	Proof of Authority
PoE	Proof of Engagement
QML	Quantum Machine Learning
RF	Random Forest
RISs	Reconfigurable Intelligent Surfaces
SDN	Software Defined Networks
SHAP	SHapley Additive exPlanations
SIM	Subscriber Identification Module
SIMs	Stacked Intelligent Surfaces
SIEM	Security Information and Event Management
SOAR	Security Orchestration, Automation, and Response
SPoF	Single Point of Failure
SVM	Support Vector Machine
TCP/IP	Transmission Control Protocol/Internet Protocol
TF-IDF	Text Frequency-Inverse Document Frequency
TL	Transfer Learning
VR	Virtual Reality
XAI	Explainable Artificial Intelligence
XR	Extend Reality
5G	Fifth Generation Network
6G	Sixth Generation Network

C. Related Surveys in Metaverse Security

The security challenges and risks within the Metaverse converge a vast array of cyber threats that have gradually gained attention as bad actors exploit personal information, user behavior, and communication links. In this section, we provide an overview of existing surveys that have explored security and privacy challenges within the Metaverse. By comparing these studies, we aim to highlight the specific focus areas of each, identify gaps in the current literature, and highlight the unique contributions of this article. This comparison will establish a clear context for understanding how our work advances the field of Metaverse security, particularly in underexplored areas.

Huang et al. [14] reported various security and privacy threats in the Metaverse and proposed a comprehensive security framework. They discussed risks such as personal information leakage, eavesdropping, unauthorized access, phishing, data injection, and broken authentication problems stemming from insecure designs during the development stages. However, their work did not explore the specific attack surfaces within the Metaverse, nor discuss the availability of datasets for security research or emphasize the suitability of NIDS models in mitigating these cyber threats.

Complementing these security concerns, Wang et al. [4] extensively examined privacy-by-design methods, particularly

TABLE II
COMPARATIVE ANALYSIS OF EXISTING METAVERSE SECURITY SURVEYS AND *MetaWatch*

Authors	Date	Focus	Taxonomy	Attacks	Attack Surfaces	Datasets	Metaverse NIDS
[15]	2021	Discussed the foundations of the Metaverse, and highlighted privacy & security issues	✓	social engineering, avatar cloning	✗	✗	✗
[4]	2022	Examined privacy-by-design data sharing, ethical concerns and digital footprint protection methods in the Metaverse	✓	social, governance, network, authentication and network-related threats	✗	✗	✗
[20]	2022	Categorized Metaverse security threats into: identity related threats, intellectual property violation, and malicious VR-quality rendering	✓	governance, smart contracts, impersonation, identity, anonymity, malicious authentication and data poisoning attacks	✗	✗	✗
[16]	2022	Advocated for vigorous laws and regulations to govern Metaverse security	✗	social engineering DDoS, shoulder-surfing, malicious authentication	✗	✗	✗
[22]–[24]	2022	Generically highlighted security threats and challenges within Metaverse	✗	data poisoning, malicious authentication, identity, biometric, cyberstalking, DDoS,	✗	✗	✗
[18], [19]	2023	Examined security and privacy concerns of the Metaverse philosophically	✗	data poisoning attacks, consensus/smart contracts vulnerabilities, insecure APIs	✗	✗	✗
[21]	2023	Examined open security challenges with respect to data heterogeneity, authentication and intrusion detection in 5-6G-enabled Metaverse	✓	semantic. man-in-the-middle, inference, identity-based attacks	✗	✗	✗
[3]	2023	Emphasized on the need to develop explainable Metaverse security models for fairness	✓	SPoF-related attacks, algorithm biases	✗	✗	✗
[14]	2023	Summarized security/privacy risks in the Metaverse owing to insecure design stages	✓	social, immersive, real-world, and data poisoning attacks	✗	✗	✗
This survey	N/A	This paper provides a summarized overview of Metaverse security challenges, and a taxonomy of NIDS methods, challenges and future directions towards a secure Metaverse.	✓	social engineering data poisoning, identity-based, network-based,	✓	✓	✓

data sharing, ethical considerations, and digital footprint protection of users for safe and beneficial Metaverse interactions. While they discussed data availability threats such as Single Point of Failure (SPoF), DDoS, and Sybil attacks, their study also did not address AI-enabled intrusion defense and detection methods that could aid in developing effective security solutions for the Metaverse and Metaverse NIDS. Similarly, Di Pietro and Cresci [15] discussed the foundations of the Metaverse and its enabling technologies, highlighting privacy and security issues introduced for users and stakeholders, but without elucidating on the technical specifics of attack surfaces, the critical role of datasets, and the relevance of Metaverse NIDS.

Additionally, Chen et al. [16] advocated for vigorous laws and regulations that would foster authentication methods, anonymity, and accountability issues that hindered Metaverse threat remediation. They argued that the Metaverse would be difficult to govern, especially with the lack of cross-border cooperation, increased cyber threats in the *darkverse*, and other privacy issues challenging to control. Despite this, their study did not examine comprehensive defense methods and approaches like ours. Similar survey works in [18] and [19] examined the security and privacy concerns of the Metaverse philosophically. In [18], the buckets effect was proposed and aimed at rethinking Metaverse security issues while highlighting the need to protect user information, communication channels, scenario applications, and goods, i.e., tamper-resistant ownership support. However, these surveys also did not provide a technical analysis of attack surfaces, intrusion detection methods, or the concept of Metaverse NIDS.

Furthermore, Ali et al. [20] summarized privacy and security threats categorized as follows: identity-related threats, intellectual property violations, (such as data tampering), malicious VR-quality rendering, as well as SPoF, DDoS, and Sybil attacks previously highlighted in [4]. Their work, like others, did not address Metaverse NIDS, the attack surfaces, or the datasets available for Metaverse security research. A recent review by Adil et al. [21] examined open security challenges within 5-6G-enabled Metaverse systems anticipated by 2028. While they advised on the feasibility of electronic Subscriber Identity Module (e-SIM) cards for efficient security verification and suggested using ML-enabled algorithms to detect heterogeneous network traffic anomalies, their study did not consider Metaverse NIDS, attack surfaces, or datasets. Similarly, Rahman et al. [3] explored the advancements and challenges of trustworthy and secure Metaverse ecosystems, emphasizing the need for XAI security models, yet without

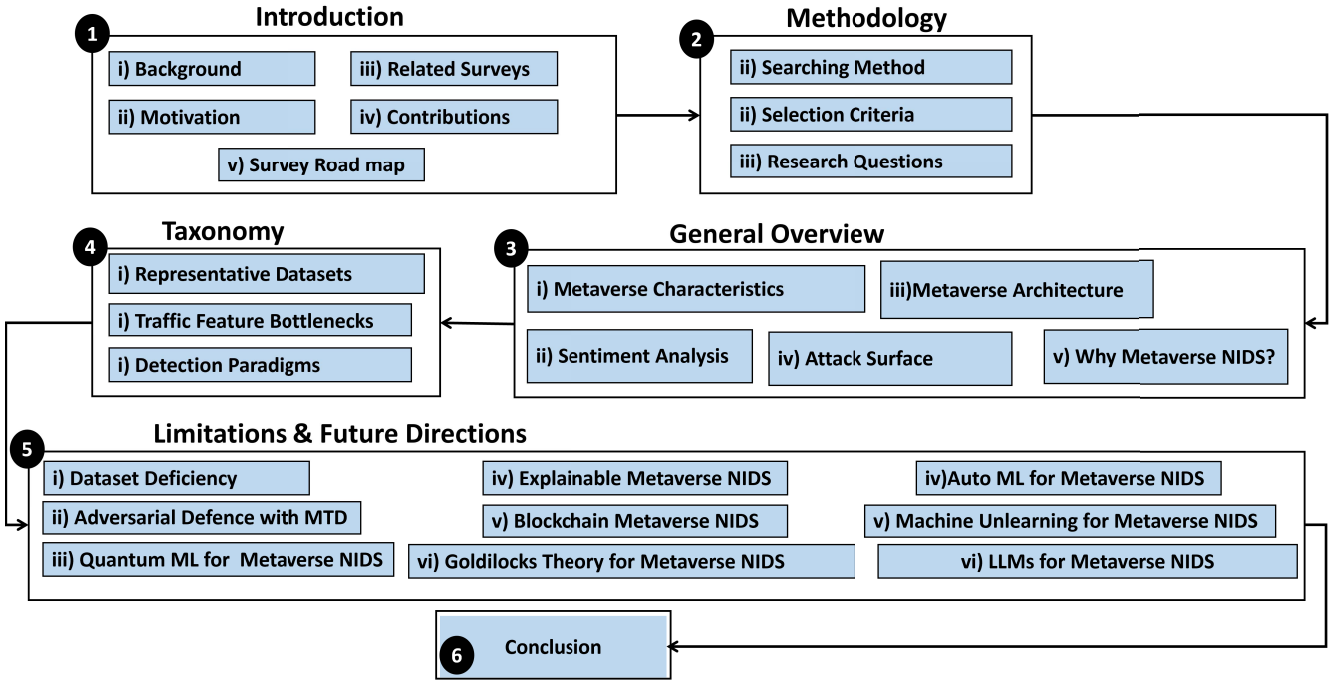

Fig. 1. MetaWatch survey roadmap—a structured overview highlighting the survey's introduction, general overview, taxonomy, limitations, and future directions.

discussing NIDS models in the Metaverse, the attack surfaces or the datasets crucial for developing these models.

Together, these studies form a cohesive narrative on the current and future privacy and security challenges in the Metaverse. However, they largely overlook technical aspects like attack surfaces, dataset availability, and a comprehensive taxonomy of intrusion detection methods essential for advancing Metaverse security.

In contrast, our survey explicitly addresses these gaps by providing a detailed review of the attack surfaces within the Metaverse, discussing the datasets available for Metaverse security research and presenting a taxonomy of Metaverse NIDS. As summarized in Table II, our study not only reviews existing NIDS frameworks but also highlights the importance of these underexplored areas. Our core aim is to provide a comprehensive analysis of Metaverse NIDS frameworks, summarize their challenges and limitations, and offer potential solutions for future research, thereby contributing uniquely to the domain of Metaverse security.

D. Contributions

This study's contributions can be summarized as follows.

1) We systematically analyze previous studies within Metaverse security and their unique contributions.
2) We analyze the research gap surrounding user sentiments of Metaverse security, Metaverse characteristics, and current attack surface employed by attackers to guide appropriate defense.
3) We present a taxonomy of Metaverse NIDS specifically, dissecting adopted datasets, network traffic feature bottlenecks, and detection algorithms to guide interested researchers and Metaverse stakeholders in this domain.
4) We discuss future methods and directions worth exploring toward Metaverse NIDS.

E. Survey Road Map

As illustrated in Fig. 1, this section provides a road map of this study. Section I introduces this study's background, motivation, and specific contributions. Section II lays out the methods employed for the survey article. Section III furnishes readers with an overview of Metaverse security, especially its characteristics, user sentiments, surface attack mechanism, and the need for Metaverse NIDS. Section IV focuses mainly on the results gathered from this review. First, we provide a taxonomy of Metaverse NIDS. Next, we review the employed datasets for Metaverse NIDS. Also, we summarize traffic feature challenges and how previous literature has addressed them. Furthermore, we provide the detection paradigms in the domain of Metaverse NIDS, which categorizes the strengths and key features of training models, methods, and algorithms. Section V summarizes the limitations of prevailing Metaverse NIDS methods and suggests future directions that can foster robust defense, trustworthiness, and efficiency for Metaverse NIDS. Finally, this article addresses some limitations encountered and draws to a close in Section VI.

II. Methodology

In this section, we provide a systematic description of the reviewing methodologies employed in this study, drawing inspiration from the PRISMA meta-analysis guidelines (Moher et al. [25] and Njoku et al. [26]) and the mente-facto conceptual design methodology [27]. By adopting these methodological approaches, articles published between

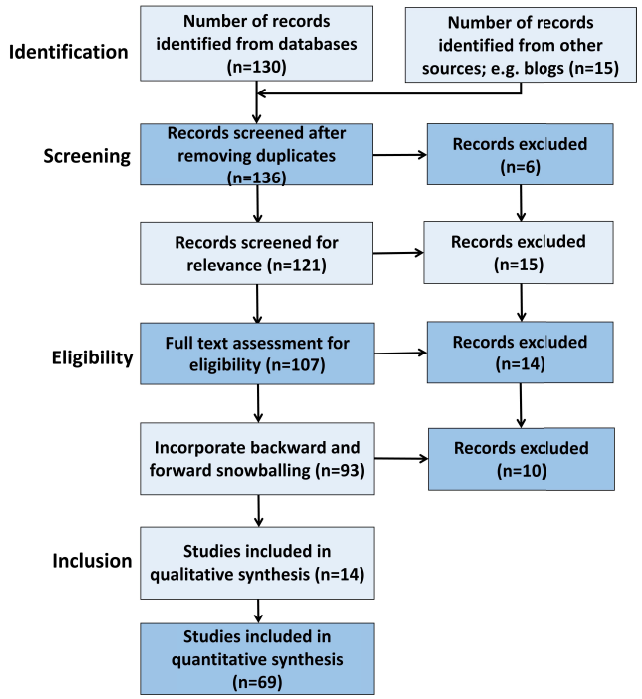

Fig. 2. Illustration of a methodological approach using PRISMA guidelines and snowballing for reviewing 65 papers, focusing on 14 studies on NIDS in the metaverse.

2021 and 2024 were considered recent and state-of-the-art. However, we recognize the importance of historical perspectives; therefore, the publication year was deemed irrelevant in specific contexts. To comprehensively identify relevant studies in the field of computer science and engineering, we conducted searches in various databases, including IEEE Xplore, arXiv, ScienceDirect, Wiley, Springer, MDPI, Taylor & Francis, as well as academic platforms such as Academia, ResearchGate, Sage, and Google Scholar.

Additionally, scholarly applications like Researchrabbit and Consensus.app were utilized to gather related materials for our study. To capture the most up-to-date literature, we incorporated the snowballing method. We systematically examined the references of included studies (backward snowballing) and the citations of included studies (forward snowballing) to identify newer studies that may not have been captured through our initial database searches. Table III provides a frequency allotment of the sources of surveyed publications, highlighting the databases with the most articles in the target domain.

A. Article Searching Method

The studies used for this study were obtained from prominent databases and nondatabase sources. These studies were retrieved from a total of 8 databases, including: *Taylor & Francis*, *ScienceDirect*, *Springer*, *MDPI*, *Wiley*, *Frontiers*, *Hindawi*, and *IEEE Xplore*. The databases were searched using keywords like *Metaverse Intrusion detection*, and *Metaverse Security*, or *VR Intrusion detection*, or *Digital Twin Intrusion Detection* and *Metaverse Review* or *Metaverse Security review*, or *Metaverse challenges* and *Digital twin Challenges* and *Review* or *Metaverse Cybersecurity*.

As illustrated in Fig. 2, studies were screened using two approaches (PRISMA and Snowballing); based on relevant databases and using other methods. A total of 112 records were identified after screening for duplicate records and other reasons from relevant databases. 4 records were not retrievable, leaving 108 for screening. These records were then screened for relevance. 20 records that failed to meet the keywords requirement were excluded, leaving a total of 88 records. These records were further assessed for eligibility. 6 records were excluded using the predefined inclusion criteria, leaving 82 records. Some records were excluded due to a lack of depth or similar content. Ultimately, 65 records were used in this review; however, 14 records were used for quantitative analysis, while 51 records were used for qualitative analysis.

B. Selection Criteria for Study

This article's inclusion and exclusion criteria for selection included the following.

1) All original articles published in journals and conference proceedings were considered.
2) All blogs and organization websites that contain detailed information within the scope of the study and cover the case studies addressed were given consideration.
3) All studies that reviewed Metaverse security issues were considered for the qualitative analysis (a total of 51).
4) All studies that reviewed the studies that have applied NIDS for Metaverse security were considered for quantitative analysis (a total of 14. See Fig. 9).
5) All articles and website posts must have been written entirely in English.
6) All papers with access restrictions could not be retrieved and thus excluded.
7) All relevant studies that fall within the reporting years of 2021 to 2024 are considered, with the year 2021 marking a significant period during which the Metaverse experienced widespread progression, investments, and severe cyberattacks [3]. Our motivation for selecting the year 2021 as a starting point is based on the fact that it marked the beginning of extensive research into the exploration of the Metaverse and its security vulnerabilities [26]. Furthermore, 2021 was notable for generating considerable excitement, product releases, and research within the Metaverse, highlighted by Facebook's rebranding to Meta.[1] Therefore, the last 4 years of Metaverse security studies were considered in this study.

C. Research Questions

The following research questions are formulated to guide this study coined as-*MetaWatch*.

1) Research Question 1 (RQ1): What are the prevailing user sentiments regarding security and privacy concerns within the Metaverse?

[1]https://www.forbes.com/sites/bernardmarr/2022/03/21/a-short-history-of-the-metaverse/?sh=332e44f15968

TABLE III
PRESENTATION OF ARTICLE SOURCES USED IN THIS STUDY

Database Source	IEEE Xplore	ScienceDirect	MDPI	Wiley	Springer	Frontiers	arXiv	Other Sources	**Total**
No. of documents	43	8	3	1	1	2	4	21	83 papers
Percentage (%)	51.80	9.63	3.61	1.20	1.20	2.4	4.81	25.30	100(%)

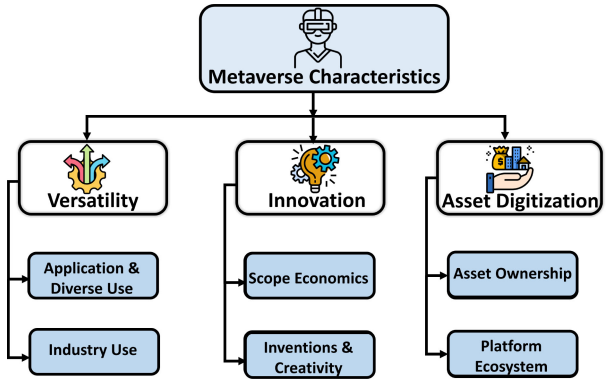

Fig. 3. Categorization of metaverse characteristics in terms of *versatility, innovation, and asset digitization.*

TABLE IV
SUMMARY OF METAVERSE STANDARDS

Standard	**Focus**
ISO/IEC 23005	Multimedia creation, support, and consumption in the Metaverse
IEEE 2888	VR interoperability, communication and interactions
IEEE P1589	Defines standards for networked virtual environments for seamless interaction
IEEE P2048	Focuses on data formats and representation in the Metaverse
IEEE P7016	Ethical concerns in Metaverse design
ISO/IEC JTC 1/SC 41/WG6	Focuses on Digital Twin (DT) standardization

2) Research Question 2 (RQ2): What are the prevailing challenges and vulnerabilities in the Metaverse that necessitate the development of AI-enabled NIDSs?

3) Research Question 3 (RQ3): How do existing studies address the challenges and limitations of Metaverse NIDS regarding dataset acquisition, network traffic feature dimensionality, training methods and detection algorithm design?

4) Research Question 4 (RQ4): What are the key characteristics and unique features of AI-enabled Metaverse NIDS that distinguish them from traditional NIDS solutions?

5) Research Question 5 (RQ5): What are the future directions and potential advancements in the development of trustworthy and efficient Metaverse NIDS?

III. GENERAL OVERVIEW

Before diving into Metaverse NIDS, this section discusses the core characteristics of the Metaverse and its attack surface.

A. Metaverse Characteristics

As introduced by early literature [4], [24], the core characteristics of the Metaverse, as categorized in Fig. 3, revolve around a sense of *immersiveness and interconnectedness*, where users can collaborate seamlessly with both virtual and real objects. Neal Stephenson coined the term Metaverse to describe a collective virtual open space where users interact through digital avatars [1]. The Metaverse is characterized by its *versatility and innovation*, and it also offers users the privilege of *asset digitization*. Due to its versatile nature that incorporates visual, spatial and experiential elements, the Metaverse has found widespread applications in various sectors, including healthcare, real estate, smart cities, games and entertainment, military and industry [3], [4].

To manage the diversity of definitions and Metaverse architectures, various standards have been proposed, including ISO/IEC 23005, IEEE 2888, IEEE P1589, IEEE P2048, and IEEE P7016, as summarized by Ali et al. [20] and Rawat and Alami [28]. Each of these standards individually addresses interface requirements, data formats, marketplace creation, definitions and terminologies, ethical considerations in Metaverse design, and digital twins, as outlined in Table IV.

B. Sentiment Analysis

This part of our study substantiates existing cybersecurity concerns within the Metaverse. It focuses on evaluating the effectiveness of the prevailing user sentiments regarding security within Metaverse and the research question posed earlier in *(RQ1)*. First, a wide range of users have shown serious concerns about the security of the Metaverse [3], [4]. In this part of this article, we bridge the research gap regarding Metaverse-related perceptions and sentiment by employing a sentiment analysis experiment to scrap tweets (opinions) and evaluate how pessimistic, optimistic, or neutral users perceive Metaverse security (main keyword).

Furthermore, a limited academic license from Twitter was obtained to scrape over 1000 relevant tweets that mentioned Metaverse or related terms from January to March 2023. Next, the tweepy and textblob libraries in Python are used to evaluate each tweet's sentiment polarity and subjectivity and add these columns to the data frame. A preprocessing was done to remove Web links, emojis, hashtags, usernames, white spaces, and nonalphanumeric characters. Subsequently, the text data is vectorized using the Text Frequency-Inverse Document Frequency (TF-IDF) method, which assigns weights to the words based on their frequency and importance.

To assess the sentiment classification performance, we evaluated the classifier using accuracy as the primary metric. A fast and accurate ML algorithm, the light gradient boost classifier, is utilized to classify the TF-IDF weighted tweets into positive, negative, or neutral sentiments. The classifier

achieved an accuracy of 97.39%, indicating high performance and reliability. The results revealed a higher proportion of positive sentiments than neutral and negative sentiments surrounding the scraped tweets.

The sentiment results presented in the GitHub repository[2] suggest that many users may have a favorable perception of their security in the virtual world but also express severe concerns and uncertainties. The closely analyzed sentiments in the CSV file provide more valuable insight into high expectations of Metaverse security, justifying this study's need for in-depth investigations and further research toward stronger security solutions for the Metaverse. Full code experiment with the dataset is provided in the GitHub repository.

Addressing RQ1

RQ1: What are the prevailing user sentiments regarding security and privacy concerns within the Metaverse?

To directly address and emphasize RQ1, sentiment analysis was conducted on a dataset of 1082 Metaverse-related tweets. Following preprocessing with TF-IDF and classification via TextBlob, scraped tweets were categorized into positive, neutral, and negative sentiment classes. Notably, negative sentiment reflects persistent user distrust in the resilience of Metaverse platforms against immersive and emerging cyber threats.

Our findings confirm that security and privacy remain central to the confidence and acceptance of Metaverse users. The demonstrated security concerns further justify the need for this study's investigation into advanced privacy-preserving methods and collaborative NIDS in the Metaverse [11].

C. Metaverse Architecture

The overall architecture of the Metaverse, as supported by extant literature, Chen et al. [16] comprised a *physical world and a virtual world* with its key technologies. The architecture of the Metaverse, as shown in Fig. 4, is structured from the bottom up, with each layer building upon the previous one to create a fully immersive virtual environment.

The *physical world* (resource pool layer) forms the foundation, consisting of essential infrastructure like humans, IoT devices, 5G/6G networks, edge nodes, and cloud computing. These elements provide the computational power and connectivity necessary for the Metaverse. Above this is the *key enablers* layer, which transforms the raw resources into more intelligent systems. These technologies enable real-time processing, adaptive learning, and dynamic simulations, ensuring the system's responsiveness and intelligence. The topmost *virtual world* layer results from the lower layers' integration, delivering immersive experiences such as AR, VR, MR, XR, and advanced digital twins. This layer represents the interactive and immersive output that defines the Metaverse. The relationship between the components of the *physical world* and the key enablers, which together give rise to the virtual world, is detailed in Fig. 4 to enhance understanding of the Metaverse architecture.

Humans/Users: First, the Metaverse will not exist without humans or users. Users within the Metaverse are responsible for exchanging sensory, radar, and localization data for immersive experiences, enabling real-time interactions and personalized content delivery in the Metaverse [4].

IoT: The Metaverse is built on IoT devices that record activities, expressions, and interactions of the physical world with various sensor technologies, thus enhancing the interaction and integration between the physical and digital worlds of the Metaverse [11]. The Metaverse integrates IoT technology as a bridge between real and virtual objects. Most head-mounted displays (HMDs), such as Oculus, Google Glass, and HoloLens, use embedded IoT sensors to exchange sensory, biometric, and location data. Network connections from HMDs to host computers are established using the transmission control protocol/Internet protocol (TCP/IP) (the networking infrastructure that allows IoT devices to connect, communicate, and share data over the Internet) and Wi-Fi connections for data transport [29].

5G/6G Networks: High-speed Networks are essential for real-time communication, data transfer, and interaction between users and haptic devices. While 5G technology has witnessed commendable advancement over previous generations, its limitations affect its performance within the Metaverse. 5G offers latency as low as one millisecond and higher throughput than 4G; however, this presents challenges for the ultralow latency and high bandwidth demands required for fully immersive and interactive Metaverse experiences.

While improved, the throughput limitations of 5G are also a constraint when considering the massive data exchange needed for real-time communication and high-definition (HD) content streaming across various devices and users in the Metaverse [4]. 6G technology, on the other hand, promises to overcome these limitations by utilizing wireless data transmission over the terahertz (THz) spectrum. This advancement significantly reduces latency, achieving ultralow delays of around 100 microseconds, approximately 1000 times lower than 5G. Moreover, 6G offers enhanced throughput, potentially reaching up to 500 Gb/s (a case study of Japan's DOCOMO firm) compared to 5G [30].

Recent explorations of reconfigurable intelligent surface (RIS) and stacked intelligent metasurfaces (SIMs) technology in 6G further improve network efficiency, coverage, and security, providing dynamic control over the electromagnetic environment to optimize signal propagation. These enhancements make 6G suitable for the Metaverse, where real-time responsiveness, seamless connectivity, and secure communication are paramount. SIMs [31], compared to RISs [30], push the boundaries of wireless communication and 6G for Metaverse networks. They are stacked with ultrathin programmable layers, just like artificial neural networks. They are proposed to dynamically propagate electromagnetic

[2] *https://github.com/nkoro/MetaverseTwitterSentiments/tree/main*

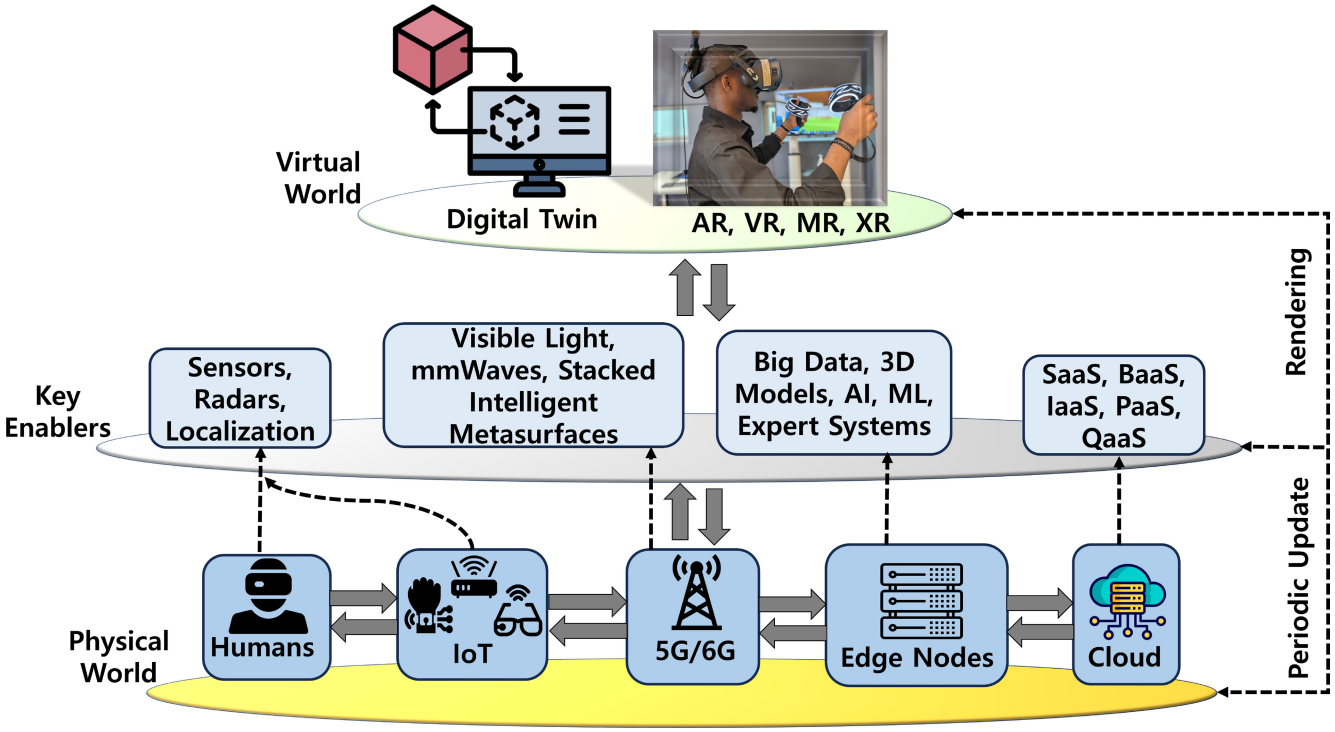

Fig. 4. Visualization of the metaverse architecture, highlighting IoT, 5G/6G, edge nodes, and cloud computing as its foundation and enablers.

waves at the speed of light while reducing hardware and energy costs, thereby attaining the requirements for low latency, communication/spectrum efficiency, high throughput, and security, critical for realizing the full potential of the Metaverse.

This review focuses on an integrated 5G and 6G network utility to support the Metaverse's demands [21]. While 5G offers foundational improvements, its limitations in latency and throughput make it less suitable for high-demand applications. However, 6G is highly prioritized, especially in areas requiring ultralow latency, high throughput, and enhanced security. Including SIMs in 6G further supports this preference, ensuring efficient, real-time, and secure communication critical for the Metaverse's full potential [32], [33].

Edge Nodes: To mitigate issues related to redundant network traffic, high latency, and quality of service/experience in the Metaverse, multiaccess edge computing architectures have become a key requirement to bring computing servers closer to data sources [34]. In vehicular Metaverses, for example, edge servers are connected to vehicles, users, and infrastructure to collect, exchange, and render simulated vehicular information with low latency [35], [36]. As shown in Fig. 4, edge nodes are facilitated with key enabling technologies like big data, simulated 3-D models, AI, ML, and expert systems to facilitate information sharing from humans, IoT, networking, and cloud resource pools.

Cloud: Cloud computing is a foundational pillar in the Metaverse's development, providing scalable, shared responsibility pay-as-you-go computational power and flexibility required to support the realization of scalable Metaverse experiences [37]. Self-hosting Metaverse organizations accrue immense running costs and technical drawbacks during development. In contrast, key cloud services, including software as a service, platform as a service, blockchain as a service, infrastructure as a service, and quantum as a service, play crucial roles in ensuring the seamless/scalable functioning of the Metaverse without disrupting user experiences.

How These Resource Pools Also Contribute to Metaverse Intrusion Detection: The Metaverse resource pools illustrated in Fig. 4, and discussed in Section III-C, including IoT devices, 5G/6G networks, edge nodes, and cloud computing, form the backbone of the Metaverse and also significantly enhance the functionality of Metaverse NIDS as a whole, in terms of collaborative real-time detection, predictive analysis, and anomaly detection.

For instance, multiaccess edge computing supports Metaverse NIDS's objectives by positioning computational resources, like 5G/6G IoT cyber–physical systems (CPSs), closer to the data source, reducing latency and enabling faster threat detection and response. Furthermore, AI and ML technologies, powered by edge nodes and cloud computing, enhance predictive analysis within NIDS by learning from historical intrusions, anticipating potential threats, and taking proactive measures to prevent attacks [13]. Cloud services enable the execution of scalable and complex algorithms to detect anomalies in network traffic across the Metaverse. In contrast, blockchain-based cloud services can incentivize genuine NIDS participants through collaborative mechanisms [6], [11], [38]. Collectively, these technologies contribute to the development of a more dynamic, intelligent, and responsive NIDS tailored to the intricate environment of the Metaverse. Further discussion of studies that buttress the

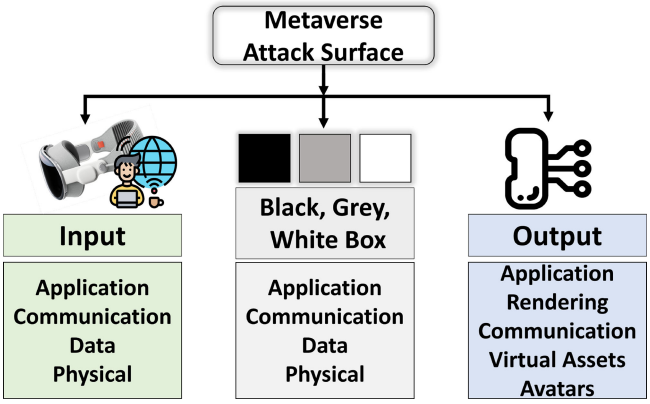

Fig. 5. Summarized illustration of the metaverse attack surface, highlighting potential exploit areas from an attacker's perspective.

utility of these technologies for Metaverse NIDS is provided in subsequent Section IV.

D. Metaverse Attack Surface

While AR, VR, and XRs in the Metaverse offer a captivating environment, cyber threats mask themselves in playful avatars or 3-D scenes while compromising existing security controls' confidentiality, integrity, and authenticity. The Metaverse attack surface is predicated on vulnerabilities propagated from regular CPSs, which attackers can leverage to harm innocent users.

The attack surface of the Metaverse, as shown in Fig. 5, can be modeled from the attacker's perspective with an input-box-output (I/O) pipeline as inspired by previous works in [10]. The input compromises all interfaces from the real world (including user data, which is the *blood* of the Metaverse, haptic devices, application services, sensors, end-user inputs, physical wires, etc.), which the attacker can manipulate. The box is a shorthand notation for a *black, gray, or white box* paradigm, which reduces difficulty levels for the attacker in each phase, respectively.

In every box scenario (black, gray, or white), the attacker needs primary information about the security configuration set (processing, computation, storage, authentication, NIDS models) to bypass defense mechanisms. Generally, black boxes are not visible or knowledgeable to engineers and developers who did not design the system. Grey boxes have parts of the system visible and modifiable to others (including users), and white boxes have the entire security configuration visible and adjustable. Output, on the other hand, can be input for another subsystem, or it may also be the final destination for the virtual world's processed data (e.g., a rendered display). A clear understanding of the I/O approach fosters proper identification of potential attack vectors and reduction of attack surfaces, as well as enabling successful Metaverse security deployments.

In addition to the Metaverse attack surface, the seven steps of the Cyber kill chain [39] applied to a wide range of CPSs, including a leading Metaverse company like *Meta* [40], enhances visibility into cyberattacks in the virtual world and increases a cyber defender's understanding of an adversary's tactics, techniques and procedures. It is called a kill chain, as illustrated in Fig. 6 because if defenders can grasp, weaken, or degrade any of the links, interrupting the attack process and improving overall security postures within the Metaverse becomes easier.

The observable trends regarding Cyberattacks in the Metaverse as compared to traditional security issues include the radicalized nature of identity theft, data breaches, unauthorized access, social engineering, and network service attacks [4], [5], [29] in the Metaverse. Specifically, nefarious hackers within these immersive platforms can use immersive and augmented user experiences to perpetuate facial, voice, and avatar behavioral cloning. Recent findings by Nair et al. [41] revealed that over 50 000 users in the Metaverse could be distinctly identified with AI by merely analyzing their head and hand motion gestures.

Compared to traditional identity theft scenarios, high-profile individuals, especially nation-states, are primary targets of defamation, impunity, and online sexual harassment in the Metaverse. Users can be hurt physically in the Metaverse with novel XR cyberattacks like *tracker attacks* (geo-locates a user directly), *chaperone attacks* (removal of safety boundaries, forcing users to hit themselves on walls or obstacles), *human joystick attacks* (metaverse zombie attack, and *overlay attacks* (ransomware, excessive light flooding, blurry or blocked vision). These novel attacks have even extended to manipulate the human brain, raising user privacy trust and techno-phobia concerns [42].

While there is limited data on the rate of cyberattacks in the Metaverse, the frequency of attacks in the Metaverse is increasing sporadically with the launch of new platforms, devices, and immersive experiences. The data presented in Table V is derived from comprehensive vulnerability records and trends obtained from the National Institute of Standards and Technology (NIST) Vulnerability Database and the Common Vulnerability Scoring System (CVSS) database. This data highlights a marked increase in the number and severity of Metaverse platform vulnerabilities from 2018 to 2024. Some affected platforms include Google Cardboard, Unreal Engine, Side Quest, Apple Vision Pro, Unity, and Roblox.

E. Why Do We Need NIDS in the Metaverse? (RQ2)

Network disruption attacks pose severe cyber risks in the Metaverse [43]. Besides understanding the objectives of adversaries and devising traditional countermeasures, the Metaverse also poses fresh cybersecurity challenges with the diversity and complexity of the technologies involving networked haptic devices, 5G/6G technologies, IoT, and blockchain [21].

Meanwhile, security threats and vulnerabilities from poorly designed Metaverse applications widen the attack surface of the Metaverse, whereby bad actors can successfully manipulate IoT-enabled haptic devices to perform malicious authentication and other related network attacks [5]. Successful compromise of IoT sensors or HMDs can modify the users' sensory, biometric, and location data, leading to identity theft, fraud, or privacy violation. Bad actors can also launch denial-of-service [7], man-in-the-room [5], or replay attacks to disrupt the *network communication* or integrity

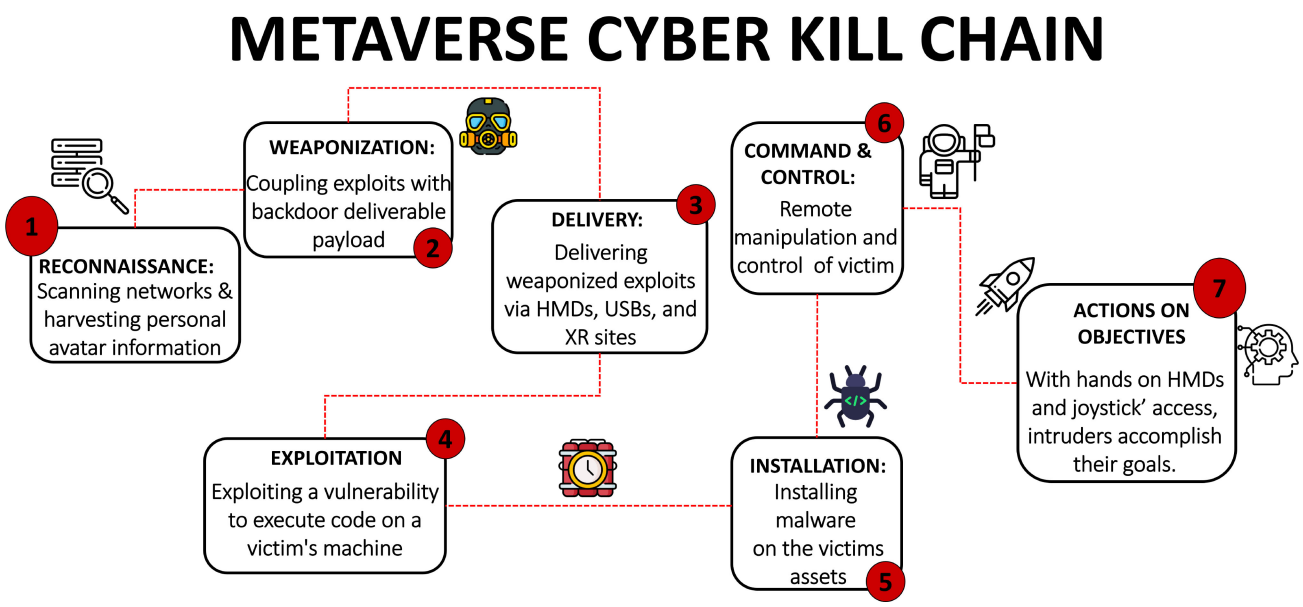

Fig. 6. Visualization of the metaverse cyber kill chain, showing the procedures and techniques attackers use to perpetrate attacks within the metaverse.

TABLE V
CHRONOLOGICAL OVERVIEW OF METAVERSE DEVELOPMENT PLATFORM ATTACKS (2018–2024)

Platform	CVE	Year	CVSS Score (010)	Description	Scale
Google Cardboard	CVE-2018-1911	2018	5.3	Sends potentially private cleartext information to the Unity 3D Stats web site	Medium
Unreal Engine	CVE-2018-10531	2019	7.5	DDoS attacks target at the 3D builders environment	High
Side Quest	CVE-2024-21625	2024	8.8	Remote code execution through malicious links in version 0.10.35	High
Apple Vision Pro	N/A	2024	N/A	Jailbreaking attack on Apple Vision Pro[3]	N/A
Unity	CVE-2024-22228	2024	7.8	XSS vulnerability attack in Dell Unity versions <5.4	High
Unity	CVE-2024-4999	2024	N/A	LigoWave devices in web-based management interfaces could allow malicious authentication attacks	N/A
Roblox	CVE-2024-31442	2024	8.8	Unauthorized admin access and logins	High

of the Metaverse, which generally affects the CIA triad of Metaverse services.

A viable security solution for monitoring and preventing network-based attacks in the Metaverse involves the design of high-performing, energy-efficient, and privacy-preserving NIDS security frameworks [11], [13], [44]. A NIDS differs from a Host-based Intrusion Detection System (HIDS) in the sense that NIDSs investigate and categorize potential threats *within the entire network,* while the Host-based IDS inspects a single device in the network. Major Metaverse corporations like *Meta* and Microsoft's *AltspaceVR* actively embrace Security Information and Event Management (SIEM), and Security Orchestration, Automation, and Response (SOAR), offering effective monitoring and detection of attack signatures and diverse network traffic behavior [45]. Robust SIEM and SOAR tools also rely on AI algorithms to categorize and detect various kinds of network traffic. Hence, utilizing AI-enabled NIDSs is essential for effective and timely detection of anomalies in the Metaverse.

Addressing RQ2

RQ2: What are the prevailing challenges and vulnerabilities in the Metaverse that necessitate the development of AI-enabled NIDSs?
In direct response to RQ2, this study discusses the highly complex attack surface within the Metaverse, stemming from the convergence of heterogeneous technologies such as IoT, vulnerable haptic systems, 5G/6G communication, and blockchain. Immersive attacks such as MitR attacks, DoS, replay attacks, and sensory manipulation, have increased volume, velocity, and variability. As such, AI-enabled NIDSs are imperative and fundamental for managing the heterogeneous nature of network traffic in the Metaverse and enabling proactive threat detection. This research question further justifies the need for a taxonomic categorization of intelligent NIDS architectures for Metaverse NIDS.

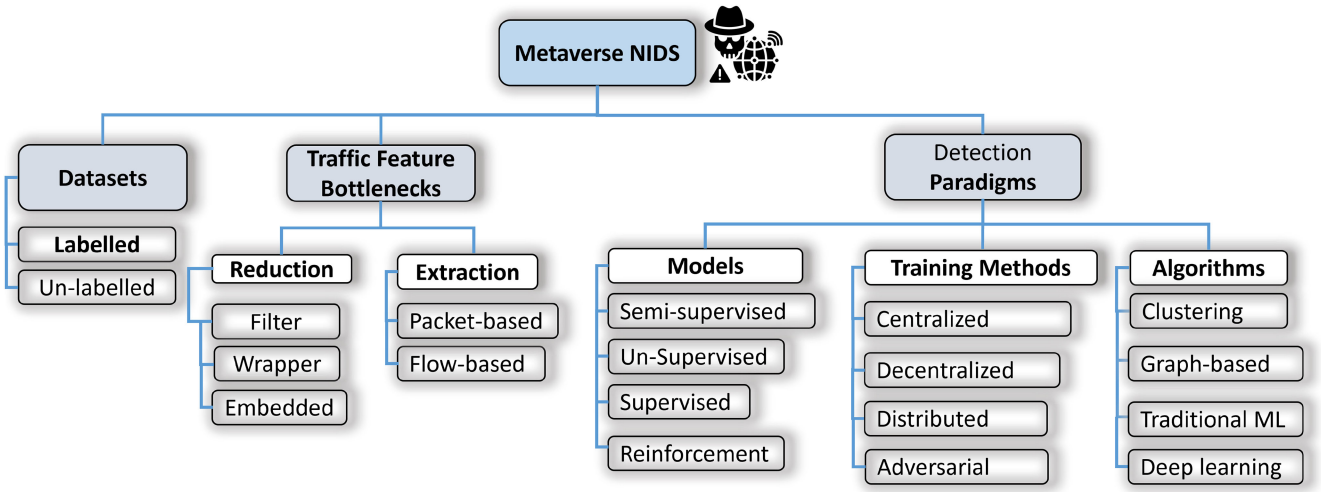

Fig. 7. Categorization of the Metaverse NIDS taxonomy with benchmark datasets, traffic feature bottlenecks, and detection paradigms.

IV. NIDS Taxonomy in the Metaverse (RQ3)

The aim of the summarized Metaverse NIDS taxonomy in Fig. 7 enables researchers to have a better understanding of the distinct methods, strengths, and limitations of previously employed methods and techniques thus directly addressing *(RQ3)*.

A. Representative Datasets and Where to Find Them

NIDS models developed specifically for the Metaverse have been evaluated on major network traffic datasets. The reviewed datasets employed for model training may be labeled or unlabelled IoT network traffic datasets. Labeled datasets contain a benign class, with other classes possessing diverse kinds of cyber attack(s) that affect Metaverse CPSs. The general justification for the use of IoT network traffic datasets may be tied to the underlying wide-range sensors and software-defined networks that Metaverse HMDs currently depend on [11]. On the other hand, image datasets have been leveraged to represent avatar facial recognition or authentication schemes [46], [47], [48]. Commonly used datasets as summarized in Table VI include the CSE-CIC-IDS2017 and CSE-CIC-IDS2018 [49], NSL KDD [50], UNSW-NB15 [51], CIC IoT 2023 [52], TON-IoT [53], BoT-IoT [54], 5G-NIDD [55], and the InSDN [56].

Salient Points: Major datasets employed for Metaverse NIDS are ostensibly realistic, as they fail to capture core Metaverse network traffic features. For example, the average bit-rate associated with VR gaming, particularly when utilizing connected HMDs and HD rendering quality, diverges significantly from that of sensors or connected IoT devices commonly employed in conventional cyber–physical dataset testbeds [57]. Concerning the high accuracy deduced from these datasets, even simple models can get near-perfect accuracy if trained and evaluated on a single dataset. The deficiency in core Metaverse network traffic datasets reduces the validity of Metaverse NIDS defenses since most datasets like NSL-KDD have become obsolete. An extension of this challenge is provided in Section V.

B. Traffic Feature Bottlenecks

Due to the interconnected and heterogeneous communication channels, the Metaverse uses to render the virtual world, its network traffic is heavily sized with high dimensionality [21]. For NIDS models to efficiently distinguish benign patterns from anomalous traffic, carefully applied feature extraction and selection techniques [13] are very useful. Feature extraction in NIDSs involves transforming raw pcap network traffic into a set of statistical features that are more meaningful and easier to analyze. Common extraction methods may include packet or flow-based methods. Some extraction tools include the CIC-flowmeter, Wireshark, Snort, tshark, tcpdump, or the NetFlow analyzer.

Feature reduction approaches on the other hand are typically undertaken by experts in network and security domains. These professionals possess both labeled datasets and a comprehensive understanding of network intricacies and algorithms, ensuring that the retained features encapsulate meaningful information for effective network analysis. Some reduction methods are the filter, wrapper, and embedded reduction techniques. By using these techniques, NIDSs can be made more efficient and effective in detecting malicious traffic in the Metaverse.

A notable cohort of studies in Metaverse NIDS have employed both the extraction and dimensionality approaches for efficient Metaverse NIDS. Truong and Le [11] and Ding et al. [12] leveraged the Autoencoder (AE) feature extraction method, to extract original network traffic features as input into the encoder network. AE typically consists of multiple hidden layers, that compress the input features into a lower-dimensional representation, known as the encoding or latent space. The decoder network, symmetrical to the encoder, takes the encoded representation and reconstructs the input features to learn patterns more intelligently unsupervised.

Nkoro et al. [13] employed a decision tree (filter-based) feature reduction method to rank network features based on their importance before supervised training. Similarly, Bütün et al. [59] utilized the Random Forest (RF) algorithm to identify the best combinations of model training features.

TABLE VI
SUMMARY OF DATASETS EMPLOYED SPECIFICALLY FOR METAVERSE NIDS AND WHERE TO FIND THEM

Dataset Owners	Name	Year	Publicly Available	Description	Studies using this dataset	Link
Canadian Institute for Cybersecurity	CSE-CIC-IDS2017 CSE-CIC-IDS2018	2017, 2018	✓	NIDS dataset for anomaly detection with over 30 features, and 7 attack scenarios.	[44], [58]	[49]
	NSL KDD	2009	✓	Improved version of the published 2009 dataset without redundant records in the train set. Contains about 311,027 Normal vs Attack classes.	[44]	[50]
	CICIoT	2023	✓	Very recent IoT cybersecurity dataset containing over 2million samples, 33 attacks, 105 devices and 7 main attack classes	[13]	[52]
Intelligent Security Group, UNSW	UNSW-NB15	2015	✓	NIDS dataset with about 7000000 samples, 48 features, and 7 cyber attacks.	[13], [32], [44]	[51]
	TON-IoT	2021	✓	NIDS dataset collected from telemetry IoT sensors, with over 23 million test/train samples, 46 features, and 10 attack classes.	[44], [59]	[53]
	BoT-IoT	2021	✓	Contains over 72,000,000 records with 5 major cyber attack classes.	[44]	[54]
UCD ASEADOS Lab	InSDN	2020	✓	Consists of 340,000 samples, 24 traffic features, and 7 cyberattack scenes.	[12]	[56]
NAU-Lincoln Joint Research Center	EdgeIIoTset	2022	✓	A comprehensive dataset for centralized and decentralized IoT NIDS, with over 20,000,000 samples, 61 features, and 15 cyberattack classes	[13]	[34]
ETRI S. Korea	5G-NIDD	2022	✓	1.2 million samples of diverse attack types in 5G networks	[6]	[55]
Dartmouth	MNIST	1999	✓	The MNIST database is a large database of 60,000 handwritten digits that is commonly used for training various image processing systems.	[60]	[47]
Roboflow Universe	MSTAR	2021	✓	The MSTAR Computer Vision dataset focuses on automatic target recognition (ATR) algorithms for synthetic aperture radar (SAR) 5,392 images	[61]	[48]

They employed an automated exhaustive grid search process that enabled a selection of network traffic features according to their importance. On the other hand, Gaber et al. [44] leveraged the kernel principal component analysis (KPCA) to reduce computational costs and real-time requirements during classification. In their experiment, the KPCA with the ToN-IoT dataset improved the detection accuracy for all attack classes compared to the results without using the KPCA.

Salient Points: In the context of Zero-day attacks, where models are less dependent on labeled datasets and may possess high dimensionality of network traffic, AEs are more capable of handling high-dimensional input features with their layering structure, and also recognizing deviations from normal traffic patterns [11]. Previous literature [9], [11] has shown that shallow ML algorithms such as RF and PCA may not be able to handle the high dimensionality/computation of network traffic, especially in the domain of Metaverse NIDS, which may have more volumes of data compared to regular datasets in IoT-based CPSs.

Additionally, the integration of ML expert/domain knowledge can yield more robust and adaptive feature extraction selection and reduction in Metaverse NIDS. Also, in cross-border scenarios, there is a need to explore feature extraction/reduction methods without compromising users' sensitive information.

Lastly, most feature extraction and reduction techniques employed for Metaverse NIDS seem to be stand-alone. There is a need to establish benchmarks and standardize evaluation metrics for comparing the effectiveness of different feature extraction and reduction techniques in Metaverse NIDS, facilitating a more unified and objective assessment of network traffic features.

C. Detection Paradigms

Metaverse NIDS core components, as illustrated in Fig. 8, comprise sensors that capture diverse traffic, management systems, audit systems, decision engines, and Graphical User Interfaces (GUIs)-based SOAR and SIEM applications [64]. Once the components of a resilient NIDS are laid, attention is shifted to the crux of AI-enabled NIDSs: the *detection paradigms*. Here, as also summarized in Table VII, a meticulous examination of Metaverse NIDS detection models, training methods, and algorithms is discussed.

1) *Detection Models:* Detection models, as provided in the taxonomy in Fig. 7, fundamentally lie in treating labeled data during the training phase. Supervised learning is overly dependent on labeled datasets, where explicit samples of normal and malicious network activities are available. Conversely, unsupervised learning thrives when labeled data is scarce or challenging to obtain, relying on the system's ability to discern patterns and anomalies autonomously. This forms a key advantage for detecting zero-day (unseen) attacks. Striking a balance between these two approaches, semi-supervised learning emerges as a pragmatic compromise. By amalgamating a limited set of labeled data with a larger pool

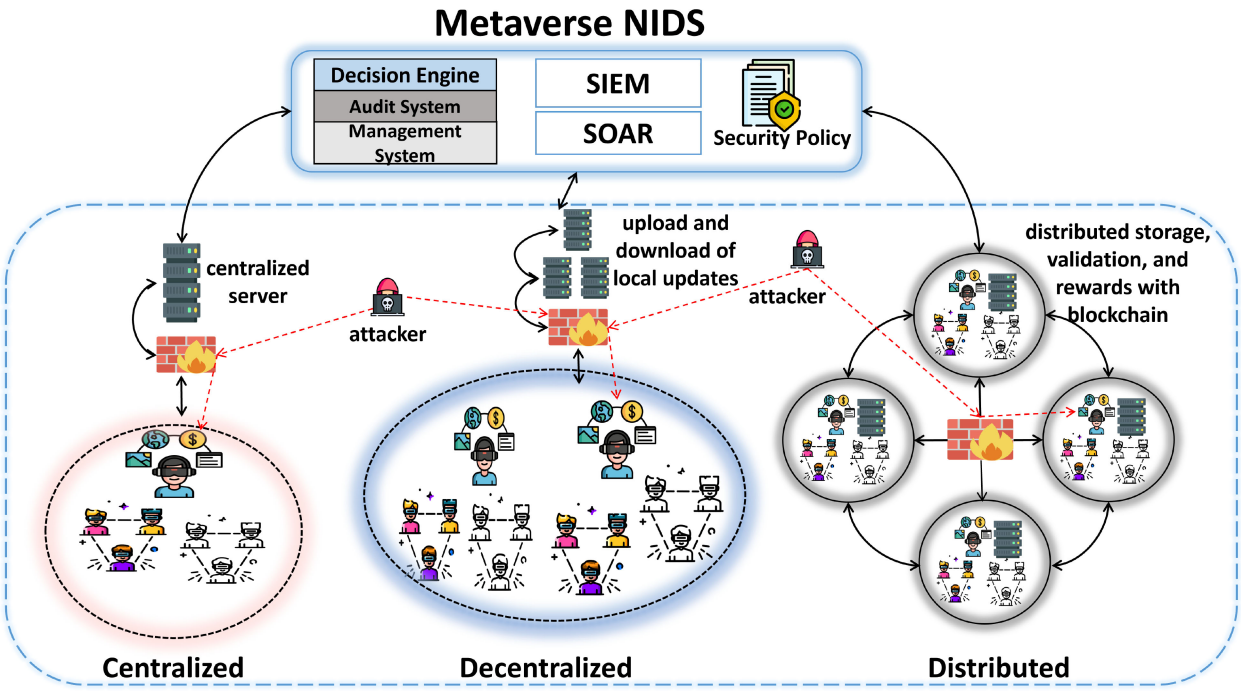

Fig. 8. Metaverse NIDS architecture. The overall system architecture relies on traffic data collection from virtual platforms, followed by threat detection using SDN-enabled intrusion detection. However, threat detection methods vary within the *centralized, decentralized, and distributed* methods.

of unlabeled data, semi-supervised learning achieves adaptability and accuracy in the face of constrained labeled NIDS datasets [11]. Furthermore, the reinforcement learning approach interconnects with the real-time network environment and employs trials to receive rewards or penalties based on the objective function of cyber threat mitigation. Reinforcement learning ultimately encourages the model to learn from dynamic network traffic to improve performance [65].

2) *Training Methods and Algorithms:* The training of NIDS models in the Metaverse is very crucial for the efficient and effective detection of cyber threats. Within surveyed literature, as illustrated in Fig. 8, the centralized, decentralized, and distributed learning methods [66] are three popular approaches employed for Metaverse NIDS model training. Each adoption approach depends on the value of the Metaverse asset being protected, which also determines the costs or computational requirements involved. A Venn diagram representing different training methods is presented in Fig. 9 for better readability.

1) Centralized Training Methods in Metaverse NIDS: Centralized training as depicted in Fig. 8 specifically involves training the NIDS model on a single server, which is useful where training data or computational requirements are limited. For example, Ding et al. [12], the early authors of Metaverse NIDS, employed a supervised and centralized training method using an RF classification algorithm to categorize anomalous network traffic in the Metaverse. To address the issue of dataset imbalance predominant in centralized architectures, they employed a GAN, to enrich the InSDN [56] training dataset and handle data imbalance problems. The GAN was employed to generate synthetic traffic patterns that closely

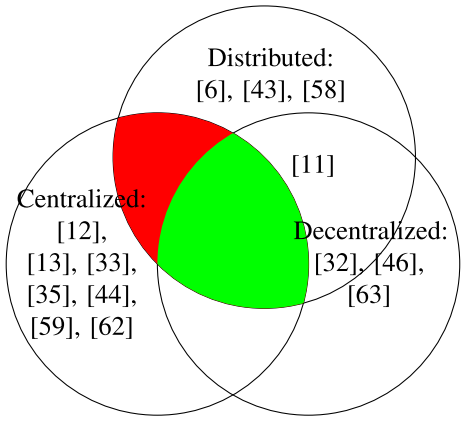

Fig. 9. Illustration of a Venn diagram representing centralized, decentralized, and distributed metaverse NIDS training approaches.

resemble real-world network behavior, with the generator, and discriminator neural network features.

The GAN can be modeled thus: Let z be the random noise vector, $G(z)$ be the generator function, and $D(x)$ be the discriminator function.

The generator's loss is represented as L_G

$$L_G = -\frac{1}{2}\mathbb{E}_z\big[\log D(G(z))\big] \tag{1}$$

while the discriminator's loss is $D(x)]$

$$L_D = -\frac{1}{2}\mathbb{E}_x\big[\log D(x)\big] - \frac{1}{2}\mathbb{E}_z\big[\log(1 - D(G(z)))\big]. \tag{2}$$

GAN's overall objective

$$\min_G \max_D (\mathcal{L}_D + \mathcal{L}_G). \tag{3}$$

TABLE VII
COMPARATIVE ANALYSIS OF METAVERSE NIDS DETECTION METHODS

Author	Dataset	FS	Model Technique	Training Method	Adversarial Training	Detection Algorithm	Evaluation Metrics	Summary
[12]	InSDN	AE	supervised	Centralized	✗	AE, and RF	RMSE, accuracy, recall, precision	Developed a GAN-based Metaverse NIDS model for zero-day attack detection, and reduce dataset imbalancing. RF is used for classification.
[44]	ToN-IoT BoN-IoT	KPCA and RF	supervised	centralized	✗	CNN	Accuracy, precision, recall, specificity, F1-score, TPR, FPR	Employed the KPCA for feature extraction and the CNN for real-time Metaverse NIDS in IoT-based SDNs.
[13]	EdgeIIoT, UNSW-NB15 CICIoT 2023	DT	supervised	centralized	✗	DNN	Accuracy, precision, recall, F1-score, CIR, DIR	Proposed a quantitative and visually interpretable DNN for Metaverse NIDS in Virtual learning platforms
[59]	ToN-IoT	RF	supervised	centralized	✗	RF	Accuracy, F1-score, TPR, FPR, TNR, FNR	Leveraged the user-plane interface for SDN-based Metaverse NIDS. They utilized the RF classifier which programmable in Intel Tofino switches.
[33]	HTTP, SMTP	-	unsupervised	centralized	✗	LSH Isolation forest classifier	F1-score, AUC Time cost	Designed a data stream anomaly detection for Metaverse healthcare NIDS, incorporating a sliding window and model update into the LSHiForest algorithm. Hash functions are leveraged to partition data streams and detect anomalies.
[61]	MSTAR	-	supervised	centralized	✓	Multiview-CNN	Accuracy, precision, recall, F1-score, FPS	Proposed an adversary detection-deactivation scheme to limit malicious participants/performances in collaborative training processes in the Metaverse.
[62]	CIFAR-10	-	un-supervised	centralized	✓	Conditional GAN	Model inversion Accuracy	Leveraged image dataset to perform a black-box contrastive projector training, model inversion and inversion optimization in Metaverse security scenarios. This security framework investigates adversarial and privacy issues that may face Metaverse NIDSs.
[63]	Ball throwing dataset	-	supervised	decentralized with FL	✗	RNN	Accuracy	Demonstrated that FL algorithms are not suitable for authentication process in Metaverse biometric authentication, due to scalability issues with an increase in number of clients and authentication modalities.
[32]	UNSW-NB15 NSLKDD	-	supervised	decentralized with FL	✗	Dynamic clustering	F1-score, AUC, recall, precision	Proposed a Metaverse NIDS framework for 6G-enabled consumer electronic devices using FL meta-learning to mitigate data imbalance and data privacy issues.
[46]	FER-2013	-	supervised	decentralized with FL	✓	CNN, RF	Attack Accuracy	Evaluated privacy defense strategies for Metaverse user security including perturbation attacks, lightweight encryption, and reconstruction attacks
[43]	Private blockchain dataset	-	supervised	distributed learning with BC	✗	Graph-based Leiden algorithm	Accuracy, F1-score, precision, specificity, execution time, storage, communication, computation	Proposed a community detection algorithm based on Leiden algorithm, to detect eclipse attacks in blockchain-enabled Metaverse.
[58]	CIC-IDS2018	-	supervised	distributed learning with BC	✗	DNN	Accuracy, precision, recall, F1-score	proposed a blockchain-based reputation system that ensures the privacy-preservation & trustworthiness of the FL process in Metaverse NIDS
[60]	MNIST	-	semi-supervised	distributed learning with BC	✓	SVM, LR	Accuracy, Attack probability.	Designed a blockchain-based privacy preserving scheme with FL, to support security and reliability of NIDSs in digital twins.
[11]	CIC-IDS2018 CIC-IDS2017 NSLKDD UNSW-NB15	AE	semi-supervised	distributed learning with BC	✓	RF	Accuracy, precision, recall, FPR, committee size, blocksize	Proposed a collaborative scheme where each node, devices and NIDS models. collaboratively trains to detect intrusions and gain rewards (NFTs).

The NIDS model proposed in this study attained an accuracy of 99.8% and 99.6% in both binary and multiclass attacks such as Normal, U2R, BFA, BOTNET, DDoS, DoS, Probe, and Web-Attack. The authors highlighted significant drawbacks, such as the erratic nature of the GAN network and its computational time constraints. Specifically, the generator and discriminator of the GAN network have difficulty reaching Nash equilibrium. Such computational limitations become a significant drawback while using GANs for Metaverse NIDS.

Bütün et al. [59] also leveraged a centralized Metaverse NIDS training mechanism to detect Metaverse cyberattacks. The user-plane interface existing in Intel Tofino switches was programmed with the real-world ToN-IoT dataset [53] to detect diverse cyberattacks in the virtual world. Their classification algorithm utilized a hyperparameter-tuned FR classifier to maximize a high detection accuracy of 99% while consuming only 5% of hardware resources and detection efficiency. Their study leaves the NIDS expert with a significant dilemma of *high attack detection accuracy or networking resource management* during the RF hyperparameter selection. For example, increasing the maximum tree depth of the RF classifier from 5 to 10 leads to a 1% gain in F1 score and a 2% drop in false positive rates but results in a 15% and 16% absolute increase in the number of required clock cycles. A drawback of this study in terms of the choice of traditional classifiers, like the RF, could be its inability to handle larger, dense, and zero-day Metaverse network traffic compared to neural networks.

Wu et al. [33] proposed DSAD, an anomaly detection scheme for 6G Metaverse healthcare, which suffers from data privacy issues and other cyberattacks such as DoS attacks. The proposed framework employs an unsupervised learning method with online learning of traffic data streams, which can capture intrinsic characteristics, infiniteness, correlation, and distribution change of the 6G Metaverse network. The attack detection phase incorporates a sliding window and model update into the LSHiForest algorithm to capture changes in network anomalies. The SMTP and HTTP datasets used in their study attained an F1 score of 72.7%, and 74.4%, respectively. The proposed model initialization of LSHiTrees, online anomaly detection, regular change detection based on normal data point distribution, and model update of anomaly detection still has some flaws in terms of *data leakage and adversarial poisoning.*

Gaber et al. [44] similarly assumed a centralized NIDS training method for IoT-based communications in the Metaverse. The proposed Metaverse NIDS framework employs a deep-learning (DL)-based convolutional neural network (CNN) that accepts the BoT-IoT [54] and ToN-IoT [53] network traffic as input to detect abnormal intrusions targeted at Metaverse CPSs. The authors first dealt with the nonlinear characteristics of high-dense Metaverse network traffic using

the KPCA feature extraction technique. Next, a CNN-based classifier yielding a 99.8% accuracy is used to classify diverse attack types. A significant drawback of the study highlighted the employment of regular CPSs and IoT datasets employed for Metaverse NIDS, which do not reflect the Metaverse networks or attacks. Their approach did not also consider adversarial defense mechanisms for the NIDS model's security.

Nkoro et al. [13] adopted a similar Metaverse NIDS framework using the EdgeIIoT [34], UNSSW-NB15 [51], and CICIoT 2023 [52] datasets to detect cyber threats that affect IoT-based communications in Metaverse E-learning platforms. The key highlight of their study revealed a high accuracy of 99% for binary classification tasks using a simple DNN, and the use of visual and quantitative XAI methods to ensure transparent and trustworthy NIDS model predictions. The visual interpretability post-hoc methods utilized were the Shapley Additive exPlanations (SHAP) and Local interpretable model-agnostic explanation (LIME) explainability methods. The quantitative XAI metrics were evaluated using calculated values of the confidence and decision impact SHAP and LIME XAI explainers. The study also debunked the adverse effects of redundant traffic features (e,g source, and destination IPs) included by previous Houda et al. [67], which culminates in misleading XAI results and biased NIDS model accuracy. Similar challenges of previous authors were highlighted in this study, which include the lack of a representative dataset for Metaverse networks, the timing constraints for post-hoc model explanations, and the lack of adversarial defense for the NIDS model.

Significant research has also been made to mitigate adversarial attacks in Metaverse NIDS solutions. Adversarial attacks in the domain of NIDS can be modeled are as follows.

Let $\vec{x_0} \in \mathbb{R}^d$ be the original network traffic sample, and C_0 be the original class. The adversarial attack $\vec{x_{adv}}$ aims to force the classification of $\vec{x_0}$ as another class C_t. The perturbed input, denoted as $\vec{x_{adv}}$, is generated to cause misclassification. The adversarial attack can be formulated as follows:

$$\vec{x_{adv}} = \vec{x_0} + \delta \tag{4}$$

where δ is the perturbation added to the original input. The objective is to find δ such that the perturbed input $\vec{x_{adv}}$ is misclassified as C_t

$$f(\vec{x_{adv}}) = C_t. \tag{5}$$

Here, f represents the classification function of the neural network classifier. Adversarial attacks may exist in white, gray, or black-box settings (where an attacker has complete, partial, or no knowledge of the model parameters), with increasing difficulty levels.

Tian et al. [62] investigated a black-box adversarial scenario that investigates training data leakage in the Metaverse. Their extensive experiments with the CIFAR-10 and CelebA datasets demonstrate the feasibility of reconstructing target images (target class) using a conditional GAN. The proposed CSMI supervised model inversion adversarial attack method is a privacy assessment tool for evaluating sensitive information vulnerabilities in the Metaverse. Although the authors leveraged an image dataset for their research, the proposed framework addresses malicious facial authentication processes that may exist in the Metaverse, as well as laying a solid foundation for adversarial vulnerability research within the Metaverse context [61].

2) Decentralized Training With Federated Learning: Primarily, the aim of the Federated Learning (FL)-enabled NIDS training method in Metaverse NIDS is to train a model *cooperatively* with the privacy preservation of edge devices, unlike the centralized approach that relies on a centralized server.

He et al. [32] employed the FL framework to categorize and detect cyber threats in 6G-enabled Metaverse environments. First, to mitigate data imbalance issues affecting traditional training methods, they employ a meta-learning scheme (meta-sampler) to obtain a balanced dataset by iteratively sampling training data. They utilize an information vector clustering algorithm for FL model aggregation and the UNSW-NB-15 and NSL-KDD datasets. The proposed approach yielded a benchmark accuracy of 83.69% while maintaining data privacy/handling data imbalance issues in Metaverse NIDS. Although the study's intended focus combines technically dense technologies like digital twins, FL, and 6G for Metaverse NIDS, the datasets employed are not reflective of core Metaverse environments, and resource costs are not considered in the study.

Adversarial attacks are also witnessed in decentralized FL methods. Sandeepa et al. [46] tackled privacy-related issues of adversarial reconstruction (perturbation) attacks in FL-enabled Metaverse security which can degrade model accuracy. They utilized a deep leakage gradient reconstruction technique to reconstruct (authentication) images, revealing that simple perturbation attacks can significantly make reconstruction attacks much cheaper for the attacker. A vital highlight of this study showed that higher model complexity plays a significant role in attack accuracy. As presented in their results, an increase in model architectures of neural networks (Basic NN, Lenet-5, Alexnet) contributed to a faster decrease in accuracy than simpler RF models. A key solution proposed in their study against adversarial perturbation attacks was the employment of differential privacy and lightweight encryption to balance security requirements in Metaverse environments.

Chen et al. [63] employed an FL method MetaGuard, for zero-trust avatar authentication in the Metaverse using biometric data. The authors leveraged the time-series features of biometric data on VR headsets to extract data with salient features for authentication. In their study, the FedAvg + FCN method achieved an accuracy of only 6.34%, while all nonprivacy-preserving models attained higher than 87% accuracy. Their results validate the tradeoff of accuracy and security claims when employing centralized versus FL privacy-preserving methods. The drawbacks of this study show the low intrusion detection performance accompanying an increase in clients using FedAvg. To tackle the decrease in accuracy issue with the increasing number of clients, the authors suggested an adaptive client selection mechanism for optimal modality

combinations for each client, with blockchain for privacy mitigation.

3) Distributed Training With Blockchain: Conventional FL frameworks rely heavily on a central server(s) for NIDS model updates, potentially leading to SPoF attacks. To address SPoF-related attacks, blockchain-assisted decentralized FL frameworks can prevent SPoF-related attacks in the Metaverse by providing distributed, secure, and transparent NIDS solutions that are resilient and robust to such frauds and attacks reliant on central authorities.

Moudoud and Cherkaoui [58] employed FL and blockchain-based logic reputation system to evaluate the reputation (reliability) of participating devices before facilitating Metaverse NIDS training in a decentralized manner. The role of blockchain in this study was very limited to storing final model updates and device reputation scores more securely. Using the CICDDoS2018 dataset [49], the FL global model attained a high 99% binary classification accuracy and satisfied the privacy requirements during training. However, this study has some drawbacks, such as the lack of blockchain experimental results to validate the resilience of model parameters against adversarial attacks and accompanying resource costs integrating blockchain with FL for Metaverse NIDS. Their study is also evaluated on a single IoT-CPS dataset, which does not reflect the core Metaverse network environment or FL scenarios.

Even blockchain-enabled Metaverse NIDS is vulnerable to cyberattacks. One of the most severe DDoS attacks launched at blockchain-enabled Metaverse environments is the eclipse attack, which encourages selfish mining and DoS attacks. Efran et al. [43] proposed a community detection algorithm (CDA) scheme using the Leiden algorithm for mitigating eclipse attacks in blockchain-enabled Metaverse NIDS. First, the study generated a blockchain dataset from the bitcoin network, including 120 nodes and simulated an eclipse attack on the network. Although the blockchain network is tested on a limited platform (bitcoin), the proposed CDA and NIDS prioritize monitoring detected attack patterns instead of analyzing the whole network. The accuracy obtained in this study reached 91.54% while preserving the limited resources on lightweight nodes of the network.

Truong and Le [11] extended the work of [32], [33], [43], and [58] in the domain of Metaverse NIDS by integrating online FL with blockchain for Metaverse NIDS -MetaCIDS. Additionally, they propose a collaborative NIDS architecture where both devices, supervised/unsupervised trainers, aggregators, and detectors receive tokens and rewards for contributing toward a secure Metaverse NIDS. The AE is used as a feature dimensionality reduction and constructed equally for online learning of zero-day network traffic. FL is employed for model weight aggregation and blockchain stores, and rewards are issued for every valid alert submitted. MetaCIDS fosters collaboration during NIDS training, addresses the problem of Zeroday attacks with online learning, and handles blockchain poisoning, inference, and SPoF concerns witnessed in the centralized approach. The semi-supervised multilayer perception model employed for diverse attack types attained a high accuracy of 99%, with benchmark metrics validating the resilience of MetaCIDS against poisoning attacks.

Zainudin et al. [6] similarly developed a collaborative incentive scheme like Truong and Le [11] using the 5G-NIDD dataset [55]. In their work, the decentralized FL-blockchain scheme, using the Proof of Authority (PoA) consensus algorithm, leveraged the Ethereum Request for Comment (ERC) tokens to minimize high transaction time costs. Within their study, a major area of interest worth exploring was the evaluation of *hybrid blockchain platforms that can further minimize transaction time in Metaverse edge devices.* Regarding model accuracy performance, the proposed method attained a 99.28% accuracy with 20 client selections. Their studies also presented blockchain-based decentralized aggregation metrics such as transaction times and gas costs. An open area of concern, as also highlighted in their studies, seeks to explore more efficient methods that can guarantee secure NIDS model exchange and high accuracy.

Salient Points: Our discussion of detection paradigms in Metaverse NIDS literature uncovered three fundamental approaches: centralized training, decentralized training with FL, and distributed training with blockchain integration. Each approach brings its unique strengths and considerations to the table.

Centralized training, exemplified by early Metaverse NIDS research [44], offers simplicity and efficiency, particularly when labeled data and computational resources are not limiting factors. However, challenges like dataset imbalance and susceptibility to adversarial/SPoF attacks necessitate innovative solutions, such as the integration of GANs [12] for data enrichment.

Decentralized training with FL emerges as a cooperative paradigm, prioritizing privacy preservation on edge devices. The meta-learning schemes and information vector clustering algorithms contribute to overcoming challenges like data imbalance while maintaining high accuracy. However, it also demands robust defenses against adversarial participants disrupting the collaborative training process [61].

Distributed training with Blockchain integration addresses SPoF concerns inherent in centralized models. By evaluating devices and intrusion alerts, using a blockchain-based logic reputation system, the approach enhances security and privacy during training. Additionally, blockchain proves valuable in storing model updates securely and mitigating most vulnerabilities encountered in decentralized FL architectures. MetaCIDS [11] takes a holistic view by integrating online FL with blockchain feature dimensionality, and more cyber attacks in both blockchain and FL approaches stand out as a promising framework for the future of Metaverse cybersecurity. However, integration of multiple advanced technologies (FL, blockchain, and the Metaverse network traffic) might introduce *complexity* in practical implementation and scalability issues in real-world IoT/Metaverse environments. Viable recommendations are provided in Section V.

TABLE VIII
PROS AND CONS OF METAVERSE NIDS ARCHITECTURES

Metaverse NIDS architecture	Pros	Cons
Centralized	High efficiency, Low communication load	High computational load, SPoF
Decentralized with FL	Medium communication/computation load and privacy aware training	Medium efficiency and minimized SPoF
Distributed with blockchain	High security, rewards mechanism, very low SPoF	High computational/communication load

Addressing RQ3

RQ3: How do existing studies address the challenges and limitations of Metaverse NIDS in acquiring datasets, dimensionality of network traffic features, training methods, and detection algorithm design?
With emphasis on the RQ3, this study reviews how Metaverse NIDS frameworks have responded to key technical barriers, including dataset limitations, high-dimensional network traffic, and preservation of privacy during training.
Central to this discourse is the comparative analysis of the centralized, decentralized (FL), and distributed methods as presented in Table VIII, each proposing distinct tradeoffs between accuracy, privacy, and scalability. The review highlights that while centralized methods demonstrate computational efficiency, they suffer from SPoF vulnerabilities and data imbalance. In contrast, FL-enabled decentralized training introduces privacy-preserving mechanisms and metalearning to mitigate data and adversarial disruptions. Distributed training frameworks that incorporate blockchain extend these gains by enhancing trust and offering incentives, but introduce system complexity and higher resource demands.

D. Distinction Between AI-Enabled Metaverse NIDS and Traditional NIDSs (RQ4)

Although NIDSs in traditional networks share similar detection algorithms and methods with Metaverse NIDS [9], there is a clear distinction in their application and efficacy due to the unique characteristics and dynamics of virtual environments within the Metaverse. The distinction we present in this section and summarized in Fig. 10, addresses the earlier posed *(RQ4)*.

First, based on the surveyed literature in Table VII, Metaverse intrusion detection adapts uniquely to the dynamic, interconnected world of VR and IoT. Furthermore, the Metaverse is significantly more data-intensive [33], [68] as compared to traditional networks in terms of streaming speed, edge computing, network data processing associated with VR gaming, and other unique characteristics of avatars. This difference in network behavior also influences the methods researchers in the Metaverse NIDS domain have employed for dimensionality reduction and detection of zero-day attacks [11], [12], [13], [44].

Another key difference between Metaverse NIDS and the traditional NIDS environments is the collaborative nature of cyberthreat detection in the Metaverse, where users, nodes, and trusted well-designed collaborators can collectively train NIDS models to keep users safe in the Metaverse while gaining rewards (NFTs, tokens, or reputation scores) for their contribution [6], [11].

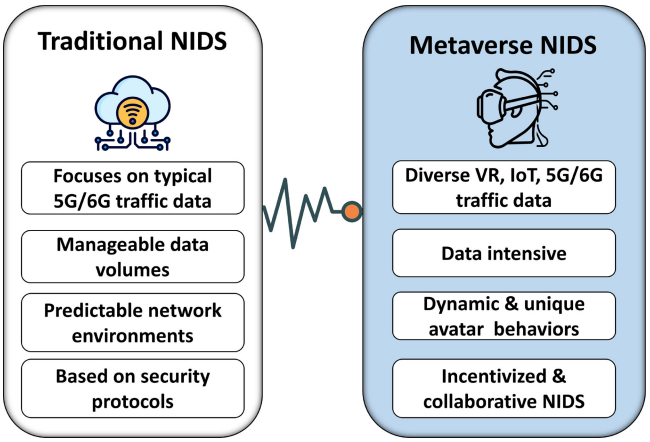

Fig. 10. Presentation of a comparison info-graph that distinguishes the traditional NIDS from metaverse NIDS.

Addressing RQ4

RQ2: What are the key characteristics and unique features of AI-enabled Metaverse NIDS that distinguish them from traditional NIDS solutions?
Based on the findings related to RQ4, this review paper articulates the fundamental distinctions between traditional NIDS frameworks and AI-enabled Metaverse NIDS as shown in Fig. 10. In summary, traditional NIDS operates within relatively static and homogeneous environments. At the same time, Metaverse NIDS must contend with dynamic, high-dimensional, and heterogeneous data streams characteristic of immersive VR, edge computing, and avatar-driven interactions.
As highlighted in our comparative analysis (Fig. 10), Metaverse environments demand enhanced scalability, data throughput, and real-time responsiveness, leading to the adoption of advanced methods for dimensionality reduction, zero-day attack detection, and collaborative learning. Moreover, Metaverse NIDS introduces novel incentive-driven trust mechanisms, such as NFTs, tokens, and logic reputation systems, to promote decentralized defense strategies.

V. Open Challenges and Future Directions in Metaverse NIDS (RQ5)

This section revisits the research question *(RQ5)* outlined at the beginning of this study, particularly focusing on *What are the open problems, future directions and potential advancements in the development of trustworthy and efficient Metaverse NIDS?* Our findings, outlined below, with contemporary literature in Metaverse NIDS, suggest pathways for future research.

A. Open Challenges

During this study, the following open issues were identified

1) *Dataset Deficiency (A Leap Forward Toward Metaverse NIDS):* The datasets used in the discussed detection paradigms, such as UNSW-NB15, NSL-KDD, BoT-IoT, ToN-IoT, and even image datasets, have undoubtedly provided valuable benchmarks for specific aspects of Metaverse security. However, it is essential to acknowledge that the Metaverse is a multifaceted space with unique network characteristics [21], including diverse communication patterns, virtual entities, and avatar behaviors. The accuracy obtained can be misleading; thus, Metaverse *NIDS needs more authentic and representative datasets*. The recent publicly available Metaverse network traffic dataset in IEEE data port [57] contains only one benign class of streamed network traffic from Metaverse platforms like Roblox and Zepeto, and is still not reflective of attack scenarios. Another challenge is that most researchers would not want to release their datasets publicly for reasons of confidentiality and privacy.
2) *Adversarial Training With MTD:* Since cyber attacks in the Metaverse have become stealthier day by day, the employment of Moving Target Defences (MTDs) can proactively offer dynamic switching of defense models to make reconnaissance attacks and other adversarial attacks in the Metaverse networks more expensive for attackers [69]. In traditional MTD networks, IP address and port mutation techniques are configured to defend against reconnaissance, eavesdropping, DDoS attacks, and other scanning-based attacks. MTDs can also apply to Metaverse NIDS.
3) *Quantum ML (QML) for Metaverse NIDS:* Traditional ML algorithms face scalability issues with large NIDS datasets (exceeding 106 106 data samples), leading to prolonged processing times and diminished accuracy, particularly in the NIDS domain. Following the Metaverse's data-intensive framework exacerbates network traffic challenges. QML has been proposed as a viable solution for NIDS within data-intensive environments like the Metaverse due to its strengths in terms of handling high-dimensional data, quantum error correction (improved and accurate detection of attacks), and quantum parallelism (accelerated training) as explored by current studies [70], [71], [72], [73].
4) *XAI for Metaverse NIDS:* The reliability and accuracy of NIDS models employed for Metaverse cybersecurity depend on the goal, truth, and context of attacks. XAI in Metaverse NIDS would specifically aim to bolster cybersecurity experts' confidence and decisions and offer explanations and reliability of predicted attacks generated by the black-box ML algorithms [74]. The summarized XAI taxonomy in Fig. 11 illustrates XAI's time, nature, and scope constraints in the domain of NIDSs.
5) *Strengthening Blockchain for Enhanced Metaverse NIDS:* Instead of overly exploiting blockchain utilities for Metaverse NIDS, a good direction worth exploring is; *what can we do for blockchain to improve Metaverse NIDS*?. Based on surveyed literature on distributed training for Metaverse NIDS [6], [11], [32], [43], [58], most proposed methods only enslave blockchain utilities for Metaverse NIDS. In contrast, blockchain smart contract codes still suffer from security vulnerabilities and computation constraints.
6) *Embracing the Goldilocks Rule for Metaverse NIDS:* The *Goldilocks rule* in the domain of Metaverse NIDS refers to *finding the optimal balance* (just the right amount) in terms of efficiency and functionality when designing NIDSs. Based on our survey, most works have employed the integration of multiple advanced technologies (blockchain, FL, IoT, edge computing), which also introduces complexity in practical implementation and scalability issues. As a result, while the models may yield high accuracies, the scalability in terms of breakout time and (usually under a minute) in large-scale Metaverse network environments needs to be extensively discussed.
7) *Auto ML for Metaverse NIDS:* Experimenting with NIDS classifiers and ML learning pipelines is a significant challenge for security experts with no technical skills and specification expertise. Continuous exploration can often be time-consuming and resource-intensive, requiring extensive manual effort to select appropriate models, tune hyperparameters, and validate performance [75]. This complexity can divert Metaverse security experts' responsibilities and attention away from critical tasks, such as detecting emerging threats and developing comprehensive security policies. *AutoML* can alleviate these challenges by automating the entire pipeline, from data preprocessing and feature engineering to model selection and optimization. With current methods like *AutoKeras, H2O.ai, and Google Cloud AutoML,* [76] Metaverse security professionals can quickly develop and deploy effective NIDS models with minimal manual intervention, freeing them to focus on higher-level security strategy and threat analysis.
8) *Machine Unlearning (MU) for Metaverse NIDS:* Growing privacy demands like the Right to be Forgotten (RTBF) within the GDPR laws [77] stipulate and require that individuals have the right to delete or revoke their private data and digital footprints at any time [78]. Within the Metaverse, if the user(s) or participating clients exercise their RTBF right, MU could dynamically remove communication patterns or traffic history data from the NIDS models. MU is envisaged to ensure that any user-specific data used to train the intrusion

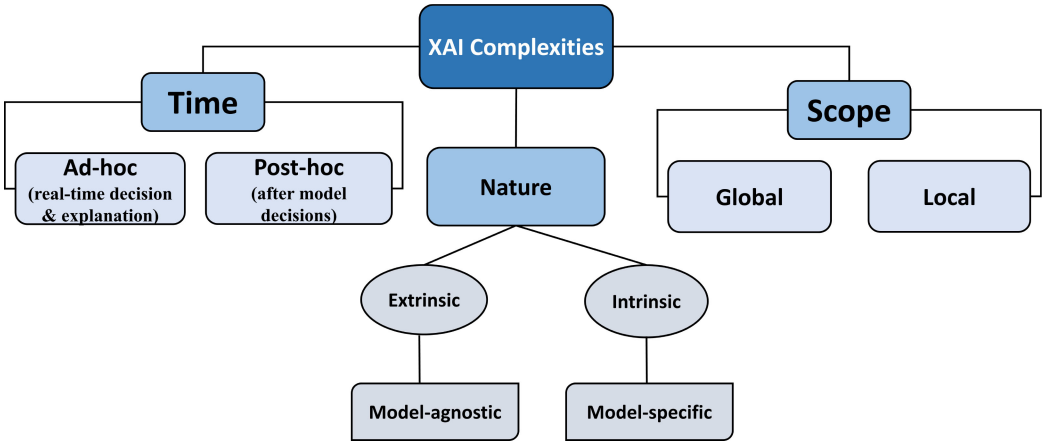

Fig. 11. XAI complexities in the domain of NIDS worth exploring. Each of these complexities presents distinct bottlenecks that require further investigation to enhance the interpretability and effectiveness of metaverse NIDS models.

detection systems or inform their algorithms is effectively forgotten without compromising the integrity and accuracy of the overall system [79]. The efficient and effective employment of *class-wise, client-wise, and sample-wise* unlearning for Metaverse NIDS remains an under-explored research area.

9) *Large Language Models (LLMs) for Metaverse NIDS:* How can NIDS models employed for the Metaverse better generalize in the presence of unlabelled heterogeneous network traffic data? This open challenge in the NIDS domain explores the effective and efficient utilization of LLMs for intrusion detection, forensics, penetration testing, and threat remediation [80]. LLMs, known for their proficiency in understanding and generating human-like text, can offer novel ways to process and interpret the vast, diverse, unlabelled traffic data typical of the Metaverse. However, the integration of LLMs into NIDS brings forth several open questions.

B. Future Directions

Future efforts towards the identified challenges in this study are outlined as follows.

1) *Addressing Dataset Deficiency:* To solve this, dataset bounty programs, competitions, or XR security boot camps can rapidly encourage the availability of realistic Metaverse NIDS datasets. A clear case study is the old NSLKDD of 2009 [50]. He et al. [9] suggested Transfer Learning (TL) techniques to challenge the lack of datasets. Still, domain mismatch and the unique characteristics of Metaverse network traffic features may be a big barrier to TL methods. A good solution calls for the simulation of a near-representative Metaverse NIDS dataset by budget-tolerant cybersecurity researchers. To solve the challenges of dataset privacy, anonymization techniques can be employed when simulating dataset test beds, which can prevent violation scenarios.
2) *Utilizing MTD for Metaverse NIDS:* Future directions worth exploring in the domain of Metaverse NIDS could involve the real-time automation of AI-enabled NIDSs to efficiently automate stable network configurations, minimal computational costs, and the following MTD objectives.
   a) What is moved? *(the network security configuration set).*
   b) When do we move? *(the timing function).*
   c) How to move *(the movement function).*
3) *QML for Metaverse NIDS:* While QML holds promise for NIDS in the Metaverse, QML still requires further research regarding hardware limitations, algorithm complexity, ethical concerns, and QML model explainability. Future directions should focus on developing more efficient quantum algorithms compatible with existing NIDS frameworks to overcome hardware constraints. Additionally, optimizing quantum error correction mechanisms will be crucial for practical implementation. Moreover, a key area for advancement will be enhancing the explainability of QML NIDS models, allowing cybersecurity experts to interpret quantum-based attack detection results transparently.
4) *XAI for Metaverse NIDS:* Current challenges in this domain worth investigating involve the complexity of model agnostic explainers like the LIME and SHAP, which cannot handle many data samples, as they can pose delayed explainability for security experts. Additionally, the rise of anomalous XAI results, which bad actors can leverage to force and fool explainers to output false confidences or explanations of NIDS predictions, thus influencing overall security in the Metaverse. For example, a poor feature selection technique, e.g., including redundant network features like source/destination IP in the training phase, leads to biased explanations of the SHAP and LIME explainers [67]. Furthermore, newer quantitative metrics like the confidence and decision impact ratios of various explainers in the domain of Metaverse NIDS as introduced by Nkoro et al. [13] can be developed further to provide cybersecurity experts more reliability and interpretation of the Metaverse NIDS model predictions.
5) *Blockchain for Metaverse NIDS:* To improve the reliability and overall security of Metaverse NIDS that

rely on blockchain, we highly recommend, as inspired by Jiang et al. [81], that improving and filtering redundant code bases of blockchain smart contracts can improve the efficiency and effectiveness of distributed Metaverse NIDS frameworks. Furthermore, as collaborative threat intelligence has gained significant interest [11] in Metaverse NIDS, malicious collaborators also try to disrupt the system to gain rewards. To solve this issue, further research in strong and secure smart contracts for issuing reputation scores should be well explored to filter out malicious trainers or collaborators in the incentive NIDS system. Novel and innovative consensus mechanisms like the Proof-of-Engagement (PoE) introduced by Nguyen et al. [82] can address the concerns of computation and scalability of distributed Metaverse NIDS.

6) *Goldilocks Rule for Metaverse NIDS:* Rethinking Metaverse NIDS design frameworks with the *Goldilocks rule* fosters the efficiency of cyberthreat detection with minimal complexities. For example, in FL environments, optimal client selection in FL-enabled Metaverse NIDS, as discussed in Section IV, faces various challenges like communication costs, fairness, heterogeneity, and resource allocation. Analyzing these challenges before training is worthy of investigation to maximize efficiency and accuracy. To this end, efficient training methods that eliminate computationally expensive nodes should be investigated to address computational costs (high gas fees incurred regardless of successful or unsuccessful transactions and training GPUs). Furthermore, *Tiny ML* methods with Quantization training can provide haptic devices with computationally friendly models that can perform basic cyber threat detection tasks like malicious authentication to keep Metaverse users safe.
7) *AutoML for Metaverse NIDS:* Future research in this domain could alleviate the concerns about the over-reliance on automated intrusion detection in the Metaverse, which could satisfy the worries of complacency or inadequate responses to sophisticated attacks since the employment of Auto ML for Metaverse NIDS also presents several challenges. One of the main concerns is the resource intensiveness in retraining models when new types of attacks emerge, as AutoML systems may not be fully equipped to handle dynamic traffic or intrusion patterns. Additionally, many AutoML solutions lack transparency in their decision-making processes, making it hard for security experts to interpret *why* a particular model flagged an intrusion, which is critical for understanding and mitigating threats, thus the need for XAI-enabled Auto ML for Metaverse NIDS.
8) *MU for Metaverse NIDS:* Future research could explore integrating proof-of-unlearning algorithms to verify that specific data of Metaverse participants has been effectively forgotten. Additionally, incorporating explainable AI (XAI) techniques into MU could enhance user trust by providing transparent explanations of the unlearning process. Another promising direction is the combination of MU with decentralized frameworks like FL to ensure that unlearning processes are secure and privacy-preserving across distributed networks. Finally, a comprehensive analysis of the overhead costs associated with unlearning, including computational and operational impacts, is essential for developing efficient and scalable MU solutions within the Metaverse [83].
9) *Efficient and Secure LLMs for Metaverse NIDS:* Integrating LLMs into Metaverse NIDS introduces several open research questions worth exploring. One such area is resource efficiency, as suggested by Fu et al. [84], who proposed adopting *lightweight* BERT LLMs. In addition to these areas, this study also suggests focusing on *LLMs' security*, especially given the dynamic nature of Metaverse networks. For example, *prompt hacking* attacks can be engineered by malicious actors to trick LLMs into generating fake or malicious output.

Addressing RQ5

RQ2: What are the future directions and potential advancements in the development of trustworthy and efficient Metaverse NIDS?

In line with the previously stated RQ5, we summarize our answers by presenting several forward-looking challenges and research gaps requiring scholarly attention to improve the robustness, scalability, and trustworthiness of AI-enabled Metaverse NIDS. For now, representative dataset scarcity remains a critical challenge in this domain. Other challenges, such as lack of model explainability, LLM security, complexities of MTD NIDS designs, and secure and efficient BC-based NIDS systems, still exist and need to be addressed. In the future, QML will offer accelerated training and computing for Metaverse NIDSs. Furthermore, MU for NIDS is underexplored but promises privacy compliance and verifiability, especially in decentralized Metaverse architectures. These future directions collectively underscore the significant motivation of this research, which charts a roadmap for the next wave of Metaverse NIDS. In light of our investigation, we conclude with the Goldilocks rule, which calls for balanced designs that are neither overly complex nor overly simplistic for Metaverse NIDS. The Goldilocks rule envisages efficiency, communication fairness, and TinyML-based models for lightweight intrusion detection in the Metaverse. Similarly, AutoML/LLM designs must evolve with transparency and adaptability to handle dynamic traffic patterns and enable secure, explainable deployment in the Metaverse.

C. Limitations of This Survey

While our systematic review provides valuable insights into Metaverse NIDS, some limitations encountered are provided for adequate consideration.

1) *Limited Existing Literature:* The parochial existing research *specifically* addressing NIDS within the

Metaverse poses a challenge. Despite the comprehensive search across databases and platforms, the number of relevant studies remains relatively small. Security researchers within the Metaverse domain could recognize this limitation and actively contribute to the field.

2) *Temporal Scope:* This study focuses on articles published between 2021 and 2024. While this ensures recent and state-of-the-art coverage, it may inadvertently exclude historical perspectives or earlier foundational work in the core domain of NIDS in CPSs and IoT.

VI. Conclusion

Metawatch surveyed all NIDS approaches in the the domain of Metaverse security. Recent attack detection designs like MetaCIDS [11] have employed a collaborative, dimensionality reduction, adversarial, and distributed NIDS management system that mitigates SPoF attacks, as well as providing incentives for collaborating nodes that jointly protect Metaverse networks against attacks. Our survey highlighted key problems, such as the nonexistence of core Metaverse NIDS datasets that allow for trustworthy evaluation of NIDS methods. Without near representative datasets in this domain, all security efforts would seemingly be *building castles on air*. To address the dataset deficiency challenge, this survey recommends new simulation experiments with anonymity/privacy designs to encourage researchers to publish Metaverse NIDS datasets freely. Regarding the trustworthiness and efficiency of Metaverse NIDS solutions, this survey highlighted the need for interpretable NIDS attack predictions to ensure more network behavior reliability by security experts. On the other hand, technically dense frameworks have the drawback of scalability. Most researchers have failed to provide the computational aspects of their integrated frameworks' detection speed. Therefore, this survey recommends a proper design of Metaverse NIDS with the *Goldilocks rule, MU, LLMs, and Auto ML* in mind to foster scalability and reliability issues while detecting cyber threats in the Metaverse.

References

[1] N. Stephenson, "Snow crash: Neal Stephenson, London, RoC(Pengiun), 1993, 440 pages," *Futures*, vol. 26, no. 7, pp. 798–800, 1994. [Online]. Available: https://doi.org/10.1016/0016-3287(94)90052-3

[2] K. A. Awan, I. U. Din, A. S. Almogren, and B.-S. Kim, "Enhancing performance and security in the metaverse: Latency reduction using trust and reputation management," *Electronics*, vol. 12, no. 15, p. 3362, 2023, [Online]. Available: https://doi.org/10.3390/electronics12153362

[3] J. R. Jim, M. T. Hosain, M. F. Mridha, M. M. Kabir, and J. Shin, "Toward trustworthy metaverse: Advancements and challenges," *IEEE Access*, vol. 11, pp. 118318–118347, 2023. [Online]. Available: https://doi.org/10.1109/ACCESS.2023.3326258

[4] Y. Wang et al., "A survey on metaverse: Fundamentals, security, and privacy," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 1, pp. 319–352, 1st Quart., 2023. [Online]. Available: https://doi.org/10.1109/comst.2022.3202047

[5] M. Vondráček, I. Baggili, P. Casey, and M. Mekni, "Rise of the metaverse's immersive virtual reality malware and the man-in-the-room attack & defenses," *Comput. Security*, vol. 127, Apr. 2023, Art. no. 102923. [Online]. Available: https://doi.org/10.1016/j.cose.2022.102923

[6] A. Zainudin, M. A. P. Putra, R. N. Alief, R. Akter, D.-S. Kim, and J.-M. Lee, "Blockchain-inspired collaborative cyber-attacks detection for securing metaverse," *IEEE Internet Things J.*, vol. 11, no. 10, pp. 18221–18236, May 2024. [Online]. Available: https://doi.org/10.1109/JIOT.2024.3364247

[7] "'Move-to-earn' application Stepn suffers cyber attack after upgrade." CoinDesk. 2022. [Online]. Available: https://www.coindesk.com/business/2022/06/06/move-to-earn-application-stepn-suffers-cyber-attack-after-upgrade/

[8] K. Dietz et al., "The missing link in network intrusion detection: Taking AI/ML research efforts to users," *IEEE Access*, vol. 12, pp. 79815–79837, 2024. [Online]. Available: https://doi.org/10.1109/ACCESS.2024.3406939

[9] K. He, D. D. Kim, and M. R. Asghar, "Adversarial machine learning for network intrusion detection systems: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 1, pp. 538–566, 1st Quart., 2023. [Online]. Available: https://doi.org/10.1109/COMST.2022.3233793

[10] J. Happa, M. Glencross, and A. Steed, "Cyber security threats and challenges in collaborative mixed-reality," *Front. ICT*, vol. 6, p. 5, Apr. 2019. [Online]. Available: https://doi.org/10.3389/fict.2019.00005

[11] V. T. Truong and L. B. Le, "MetaCIDS: Privacy-preserving collaborative intrusion detection for metaverse based on blockchain and online federated learning," *IEEE Open J. Comput. Soc.*, vol. 4, pp. 253–266, 2023. [Online]. Available: https://doi.org/10.1109/ojcs.2023.3312299

[12] S. Ding, L. Kou, and T. fa Wu, "A GAN-based intrusion detection model for 5G enabled future metaverse," *J. Special Topics Mobile Netw. Appl.*, to be published. [Online]. Available: https://doi.org/10.1007/s11036-022-02075-6

[13] E. C. Nkoro, C. I. Nwakanma, J.-M. Lee, and D.-S. Kim, "Detecting cyberthreats in metaverse learning platforms using an explainable DNN," *Internet Things*, vol. 25, Apr. 2024, Art. no. 101046. [Online]. Available: https://doi.org/10.1016/j.iot.2023.101046

[14] Y. Huang, Y. J. Li, and Z. Cai, "Security and privacy in metaverse: A comprehensive survey," *Big Data Min. Anal.*, vol. 6, no. 2, pp. 234–247, 2023. [Online]. Available: https://doi.org/10.26599/BDMA.2022.9020047

[15] R. Di Pietro and S. Cresci, "Metaverse: Security and privacy issues," in *Proc. 3rd IEEE Int. Conf. Trust, Privacy Security Intell. Syst. Appl. (TPS-ISA)*, 2021, pp. 281–288. [Online]. Available: https://doi.org/10.1109/tpsisa52974.2021.00032

[16] Z. Chen, J. Wu, W. Gan, and Z. Qi, "Metaverse security and privacy: An overview," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, 2022, pp. 2950–2959. [Online]. Available: https://doi.org/10.1109/BigData55660.2022.10021112

[17] T. A. Jaber, "Security risks of the metaverse world," *Int. J. Interact. Mobile Technol.*, vol. 16, no. 13, pp. pp. 4–14, Jul. 2022. [Online]. Available: https://doi.org/10.3991/ijim.v16i13.33187

[18] C. Zhang, X. Si, X. Zhu, and Y. Zhang, "A survey on the security of the metaverse," in *Proc. IEEE Int. Conf. Metaverse Comput., Netw. Appl. (MetaCom)*, 2023, pp. 428–432. [Online]. Available: https://doi.org/10.1109/metacom57706.2023.00082

[19] M. Tukur, J. Schneider, M. J. Househ, A. H. Dokoro, U. I. Ismail, M. Dawaki, and M. Agus, "The metaverse digital environments: A scoping review of the challenges, privacy and security issues," *Front. Big Data*, vol. 6, Nov. 2023, Art. no. 1301812. [Online]. Available: https://doi.org/10.3389/fdata.2023.1301812

[20] M. N. Ali, F. Naeem, G. Kaddoum, and E. Hossain, "Metaverse communications, networking, security, and applications: Research issues, state-of-the-art, and future directions," *IEEE Commun. Surveys Tuts.*, vol. 26, no. 2, pp. 1238–1278, 2nd Quart., 2024. [Online]. Available: https://doi.org/10.48550/arxiv.2212.13993

[21] M. Adil, H. Song, M. K. Khan, A. Farouk, and Z. Jin, "5G/6G-enabled metaverse technologies: Taxonomy, applications, and open security challenges with future research directions," 2023, *arXiv:2305.16473*.

[22] M. Pooyandeh, K.-J. Han, and I. Sohn, "Cybersecurity in the AI-based metaverse: A survey," *Appl. Sci.*, vol. 12, no. 24, 2022, Art. no. 12993. [Online]. Available: https://doi.org/10.3390/app122412993

[23] A. Gupta, H. U. Khan, S. Nazir, M. Shafiq, and M. Shabaz, "Metaverse security: Issues, challenges and a viable ZTA model," *Electronics*, vol. 12, no. 2, p. 391, 2023. [Online]. Available: https://doi.org/10.3390/electronics12020391

[24] S.-M. Park and Y.-G. Kim, "A metaverse: Taxonomy, components, applications, and open challenges," *IEEE Access*, vol. 10, pp. 4209–4251, 2022. [Online]. Available: https://doi.org/10.1109/ACCESS.2021.3140175

[25] D. Moher et al., "Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement," *Syst. Rev.*, vol. 4, no. 1, pp. 1–9, 2015.

[26] J. N. Njoku, C. I. Nwakanma, G. C. Amaizu, and D.-S. Kim, "Prospects and challenges of metaverse application in data-driven intelligent transportation systems," *IET Intell. Transp. Syst.*, vol. 17, no. 1, pp. 1–21, 2023. [Online]. Available: https://doi.org/10.1049/itr2.12252

[27] P. V. Torres-Carrin, C. S. Gonzlez-Gonzalez, S. Aciar, and G. Rodriguez-Morales, "Methodology for systematic literature review applied to engineering and education," in *Proc. IEEE Global Eng. Educ. Conf. (EDUCON)*, 2018, pp. 1364–1373.

[28] D. B. Rawat and H. El Alami, "Metaverse: Requirements, architecture, standards, status, challenges, and perspectives," *IEEE Internet Things Mag.*, vol. 6, no. 1, pp. 14–18, Mar. 2023. [Online]. Available: https://doi.org/10.1109/IOTM.001.2200258

[29] S.-Y. Kuo, F.-H. Tseng, and Y.-H. Chou, "Metaverse intrusion detection of wormhole attacks based on a novel statistical mechanism," *Future Gener. Comput. Syst.*, vol. 143, pp. 179–190, Jun. 2023. [Online]. Available: https://doi.org/10.1016/j.future.2023.01.017

[30] J. Li, S. Dang, Z. Zhang, and L. Wang, "When industrial metaverse meets 6G: The next revolution and deployment challenges," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, 2024, pp. 1–6. [Online]. Available: https://doi.org/10.1109/WCNC57260.2024.10570966

[31] X. Yao, J. An, L. Gan, M. Di Renzo, and C. Yuen, "Channel estimation for stacked intelligent metasurface-assisted wireless networks," *IEEE Wireless Commun. Lett.*, vol. 13, no. 5, pp. 1349–1353, May 2024. [Online]. Available: https://doi.org/10.1109/LWC.2024.3369874

[32] S. He, C. Du, and M. S. Hossain, "6G-enabled consumer electronics device intrusion detection with federated meta-learning and digital twins in a meta-verse environment," *IEEE Trans. Consum. Electron.*, vol. 70, no. 1, pp. 3111–3119, Feb. 2024. [Online]. Available: https://doi.org/10.1109/TCE.2023.3321846

[33] X. Wu, Y. Yang, M. Bilal, L. Qi, and X. Xu, "6G-enabled anomaly detection for metaverse healthcare analytics in Internet of Things," *IEEE J. Biomed. Health Inform.*, vol. 28, no. 11, pp. 6308–6317, Nov. 2024. [Online]. Available: https://doi.org/10.1109/JBHI.2023.3298092

[34] 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. [Online]. Available: https://doi.org/10.1109/ACCESS.2022.3165809

[35] B. D. Son et al., "Adversarial attacks and defenses in 6G network-assisted IoT systems," *IEEE Internet Things J.*, vol. 11, no. 11, pp. 19168–19187, Jun. 2024. [Online]. Available: https://doi.org/10.1109/JIOT.2024.3373808

[36] J. Yu, A. Alhilal, P. Hui, and D. H. K. Tsang, "Bi-directional digital twin and edge computing in the metaverse," *IEEE Internet Things Mag.*, vol. 7, no. 3, pp. 106–112, May 2024. [Online]. Available: https://doi.org/10.1109/IOTM.001.2300173

[37] C. Chen et al., "Privacy Computing meets metaverse: Necessity, taxonomy and challenges," *Ad Hoc Netw.*, vol. 158, May 2024, Art. no. 103457. [Online]. Available: https://doi.org/10.1016/j.adhoc.2024.103457

[38] M. Chelghoum, G. Bendiab, M. A. Labiod, M. Benmohammed, S. Shiaeles, and A. Mellouk, "Blockchain and AI for collaborative intrusion detection in 6G-enabled IoT networks," in *Proc. IEEE 25th Int. Conf. High Perform. Switching Routing (HPSR)*, 2024, pp. 179–184. [Online]. Available: https://doi.org/10.1109/HPSR62440.2024.10635989

[39] "The cyber kill chain." 2024. [Online]. Available: https://www.lockheedmartin.com/en-us/capabilities/cyber/cyber-kill-chain.html

[40] B. Nimmo and E. Hutchins, "Phase-based tactical analysis of online operations," Carnegie Endow. Int. Peace, Washington, DC, USA, Rep. 202303, 2023.

[41] V. Nair et al., "Unique identification of 50,000+ virtual reality users from head & hand motion data," in *Proc. 32nd USENIX Security Symp. (USENIX Security)*, 2023, pp. 895–910. [Online]. Available: https://doi.org/10.5555/3620237.3620288

[42] S. Qamar, Z. Anwar, and M. Afzal, "A systematic threat analysis and defense strategies for the metaverse and extended reality systems," *Comput. Security*, vol. 128, May 2023, Art. no. 103127. [Online]. Available: https://doi.org/10.1016/j.cose.2023.103127

[43] F. Erfan, M. Bellaiche, and T. Halabi, "Community detection algorithm for mitigating eclipse attacks on blockchain-enabled metaverse," in *Proc. IEEE Int. Conf. Metaverse Comput., Netw. Appl. (MetaCom)*, 2023, pp. 403–407. [Online]. Available: https://doi.org/10.1109/metacom57706.2023.00077

[44] T. Gaber, J. B. Awotunde, M. Torky, S. A. Ajagbe, M. Hammoudeh, and W. Li, "Metaverse-IDS: Deep learning-based intrusion detection system for metaverse-IoT networks," *Internet Things*, vol. 24, Dec. 2023, Art. no. 100977. [Online]. Available: https://doi.org/10.1016/j.iot.2023.100977

[45] N. Sun et al., "Cyber threat intelligence mining for proactive cybersecurity defense: A survey and new perspectives," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 3, pp. 1748–1774, 3rd Quart., 2023. [Online]. Available: https://doi.org/10.1109/COMST.2023.3273282

[46] C. Sandeepa, S. Wang, and M. Liyanage, "Privacy of the metaverse: Current issues, AI attacks, and possible solutions," in *Proc. IEEE Int. Conf. Metaverse Comput., Netw. Appl. (MetaCom)*, 2023, pp. 234–241. [Online]. Available: https://doi.org/10.1109/MetaCom57706.2023.00052

[47] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998. [Online]. Available: https://doi.org/10.1109/5.726791

[48] N. Stein, "MSTAR 2.0 Dataset," *Roboflow Universe*, vol. 86, Aug. 2021. [Online]. Available: https://universe.roboflow.com/nathan-stein/mstar-2.0

[49] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization," in *Proc. Int. Conf. Inf. Syst. Security Privacy*, 2018, pp. 1–9.

[50] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, "A detailed analysis of the KDD CUP 99 data set," in *Proc. IEEE Symp. Comput. Intell. Security Defense Appl.*, 2009, pp. 1–6. [Online]. Available: https://doi.org/10.1109/CISDA.2009.5356528

[51] N. Moustafa and J. Slay, "UNSW-NB15: A comprehensive dataset for network intrusion detection systems (UNSW-NB15 network dataset)," in *Proc. Mil. Commun. Inf. Syst. Conf. (MilCIS)*, 2015, pp. 1–6. [Online]. Available: https://doi.org/10.1109/MilCIS.2015.7348942

[52] 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. [Online]. Available: https://doi.org/10.3390/s23135941

[53] N. Moustafa, "A new distributed architecture for evaluating AI-based security systems at the edge: Network TONIoT datasets," *Sustain. Cities Soc.*, vol. 72, Sep. 2021, Art. no. 102994. [Online]. Available: https://doi.org/10.1016/j.scs.2021.102994

[54] J. Ashraf et al., "IoTBoT-IDS: A novel statistical learning-enabled botnet detection framework for protecting networks of smart cities," *Sustain. Cities Soc.*, vol. 72, Sep. 2021, Art. no. 103041. [Online]. Available: https://doi.org/10.1016/j.scs.2021.103041

[55] S. Samarakoon et al., "5G-NIDD: A comprehensive network intrusion detection dataset generated over 5G wireless network," 2022, *arXiv:2212.01298*.

[56] M. S. Elsayed, N.-A. Le-Khac, and A. D. Jurcut, "InSDN: A novel SDN intrusion dataset," *IEEE Access*, vol. 8, pp. 165263–165284, 2020. [Online]. Available: https://doi.org/10.1109/ACCESS.2020.3022633

[57] Y.-H. Choi et al., "ML-based 5G traffic generation for practical simulations using open datasets," *IEEE Commun. Mag.*, vol. 61, no. 9, pp. 130–136, Sep. 2023. [Online]. Available: https://doi.org/10.1109/MCOM.001.2200679

[58] H. Moudoud and S. Cherkaoui, "Federated learning meets blockchain to secure the metaverse," in *Proc. Int. Wireless Commun. Mobile Comput. (IWCMC)*, 2023, pp. 339–344. [Online]. Available: https://doi.org/10.1109/IWCMC58020.2023.10182956

[59] B. Bütün, A. T.-J. Akem, M. Gucciardo, and M. Fiore, "Fast detection of cyberattacks on the metaverse through user-plane inference," in *Proc. IEEE Int. Conf. Metaverse Comput., Netw. Appl. (MetaCom)*, 2023, pp. 350–354. [Online]. Available: https://doi.org/10.1109/MetaCom57706.2023.00067

[60] Z. Lv, C. Cheng, and H. Lv, "Blockchain-based decentralized learning for security in digital twins," *IEEE Internet Things J.*, vol. 10, no. 24, pp. 21479–21488, Dec. 2023. [Online]. Available: https://doi.org/10.1109/JIOT.2023.3295499

[61] P. Li, Z. Zhang, A. S. Al-Sumaiti, N. Werghi, and C. Y. Yeun, "A robust adversary detection-deactivation method for metaverse-oriented collaborative deep learning," *IEEE Sensors J.*, vol. 24, no. 14, pp. 22011–22022, Jul. 2024. [Online]. Available: https://doi.org/10.1109/JSEN.2023.3325771

[62] Z. Tian, C. Zhang, K. Sood, and S. Yu, "Inferring private data from AI models in metaverse through black-box model inversion attacks," in *Proc. IEEE Int. Conf. Metaverse Comput., Netw. Appl. (MetaCom)*, 2023, pp. 49–56. [Online]. Available: https://doi.org/10.1109/MetaCom57706.2023.00051

[63] R. Cheng, S. Chen, and B. Han, "Towards Zero-trust security for the metaverse," *IEEE Commun. Mag.*, vol. 62, no. 2, pp. 156–162, Feb. 2024.

[64] W. Li, W. Meng, and L. F. Kwok, "Surveying trust-based collaborative intrusion detection: State-of-the-art, challenges and future directions," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 1, pp. 280–305, 1st Quart., 2022. [Online]. Available: https://doi.org/10.1109/COMST.2021.3139052

[65] N. Moustafa, N. Koroniotis, M. Keshk, A. Y. Zomaya, and Z. Tari, "Explainable intrusion detection for cyber defences in the Internet of Things: Opportunities and solutions," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 3, pp. 1775–1807, 3rd Quart., 2023. [Online]. Available: https://doi.org/10.1109/COMST.2023.3280465

[66] C. Ma et al., "When federated learning meets blockchain: A new distributed learning paradigm," *IEEE Comput. Intell. Mag.*, vol. 17, no. 3, pp. 26–33, Aug. 2022. [Online]. Available: https://doi.org/10.1109/MCI.2022.3180932

[67] Z. A. E. Houda, B. Brik, and L. Khoukhi, "'Why Should I Trust Your IDS?' An explainable deep learning framework for intrusion detection systems in Internet of Things networks," *IEEE Open J. Commun. Soc.*, vol. 3, pp. 1164–1176, 2022. [Online]. Available: https://doi.org/10.1109/OJCOMS.2022.3188750

[68] Y. Cai, J. Llorca, A. M. Tulino, and A. F. Molisch, "Compute- and data-intensive networks: The key to the metaverse," in *Proc. 1st Int. Conf. 6G Netw. (6GNet)*, 2022, pp. 1–8. [Online]. Available: https://doi.org/10.1109/6GNet54646.2022.9830429

[69] X. Qin, F. Jiang, M. Cen, and R. Doss, "Hybrid cyber defense strategies using Honey-X: A survey," *Comput. Netw.*, vol. 230, Jul. 2023, Art. no. 109776. [Online]. Available: https://doi.org/10.1016/j.comnet.2023.109776

[70] E. A. Tuli, J.-M. Lee, and D.-S. Kim, "Integration of quantum technologies into metaverse: Applications, potentials, and challenges," *IEEE Access*, vol. 12, pp. 29995–30019, 2024. [Online]. Available: https://doi.org/10.1109/ACCESS.2024.3366527

[71] M. Kalinin and V. Krundyshev, "Security intrusion detection using quantum machine learning techniques," *J. Comput. Virol. Hacking Techn.*, vol. 19, no. 1, pp. 125–136, 2023. [Online]. Available: https://doi.org/10.1007/s11416-022-00435-0

[72] C. Gong, W. Guan, A. Gani, and H. Qi, "Network attack detection scheme based on variational quantum neural network," *J. Supercomput.*, vol. 78, no. 15, pp. 16876–16897, 2022. [Online]. Available: https://doi.org/10.1007/s11227-022-04542-z

[73] O. K. Nicesio, A. G. Leal, and V. L. Gava, "Quantum machine learning for network intrusion detection systems, a systematic literature review," in *Proc. IEEE 2nd Int. Conf. AI Cybersecurity (ICAIC)*, 2023, pp. 1–6. [Online]. Available: https://doi.org/10.1109/ICAIC57335.2023.10044125

[74] P. Kumar, R. Kumar, M. Aloqaily, and A. K. M. N. Islam, "Explainable AI and blockchain for metaverse: A security, and privacy perspective," *IEEE Consum. Electron. Mag.*, vol. 13, no. 3, pp. 90–97, May 2024. [Online]. Available: https://doi.org/10.1109/MCE.2023.3296222

[75] L. Yang, M. E. Rajab, A. Shami, and S. Muhaidat, "Enabling AutoML for zero-touch network security: Use-case driven analysis," *IEEE Trans. Netw. Service Manag.*, vol. 21, no. 3, pp. 3555–3582, Jun. 2024. [Online]. Available: https://doi.org/10.1109/TNSM.2024.3376631

[76] M. A. Khan, N. Iqbal, Imran, H. Jamil, and D.-H. Kim, "An optimized ensemble prediction model using AutoML based on soft voting classifier for network intrusion detection," *J. Netw. Comput. Appl.*, vol. 212, Mar. 2023, Art. no. 103560. [Online]. Available: https://doi.org/10.1016/j.jnca.2022.103560

[77] V. Ayala-Rivera and L. Pasquale, "The grace period has ended: An approach to operationalize GDPR requirements," in *Proc. IEEE 26th Int. Requirements Eng. Conf. (RE)*, 2018, pp. 136–146. [Online]. Available: https://doi.org/10.1109/RE.2018.00023

[78] C. Li et al., "An overview of machine unlearning," in *High-Confidence Computing*. Amsterdam, The Netherlands: Elsevier, 2024, Art. no. 100254. [Online]. Available: https://doi.org/10.1016/j.hcc.2024.100254

[79] Y. Yuan, B. Wang, C. Zhang, Z. Xiong, C. Li, and L. Zhu, "Toward efficient and robust federated unlearning in IoT networks," *IEEE Internet Things J.*, vol. 11, no. 12, pp. 22081–22090, Jun. 2024. [Online]. Available: https://doi.org/10.1109/JIOT.2024.3378329

[80] 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," 2024, *arXiv:2405.12750*.

[81] Z. Jiang, Z. Zheng, K. Chen, X. Luo, X. Tang, and Y. Li, "Exploring smart contract recommendation: Towards efficient blockchain development," *IEEE Trans. Services Comput.*, vol. 16, no. 3, pp. 1822–1832, May/Jun. 2023. [Online]. Available: https://doi.org/10.1109/TSC.2022.3202081

[82] C. T. Nguyen, D. T. Hoang, D. N. Nguyen, Y. Xiao, D. Niyato, and E. Dutkiewicz, "MetaShard: A novel sharding blockchain platform for metaverse applications," *IEEE Trans. Mobile Comput.*, vol. 23, no. 5, pp. 4348–4361, May 2024. [Online]. Available: https://doi.org/10.1109/TMC.2023.3290955

[83] A. Islam, H. Karimipour, T. R. Gadekallu, and Y. Zhu, "A federated unlearning-based secure management scheme to enable automation in smart consumer electronics facilitated by digital twin," *IEEE Trans. Consum. Electron.*, early access, May 3, 2024, doi: 10.1109/TCE.2024.3396723.

[84] M. Fu, P. Wang, M. Liu, Z. Zhang, and X. Zhou, "IoV-BERT-IDS: Hybrid network intrusion detection system in IoV using large language models," *IEEE Trans. Veh. Technol.*, vol. 74, no. 2, pp. 1909–1921, Feb. 2025. [Online]. Available: https://doi.org/10.1109/TVT.2024.3402366

Ebuka Chinaechetam Nkoro received the B.Eng. degree in software engineering from the Federal University of Technology, Owerri, Nigeria, in 2009, and the M.Sc. degree in IT-convergence engineering from Kumoh National Institute of Technology, Gumi, South Korea, in 2022.

He is a Research Assistant with the ICT Convergence Research Center, Kumoh National Institute of Technology, where he conducts research on AI-enabled threat intelligence and secure networked systems. In addition to his academic role, he contributes to real-world cybersecurity operations as part of Chainlabs, supporting threat analysis, anti-money laundering, and secure intelligence workflows in blockchain environments. He has also completed internships with the Nigerian Communications Commission and an early career fellowship with the Internet Society. His research interests include explainable AI for cyber–physical systems' security, secure mobility services, and threat modeling in Web3 and the industrial Metaverse.

Mr. Nkoro is an Active Reviewer for peer-reviewed journals and a member of both ISOC and the Cyber Security Experts Association of Nigeria.

Judith Nkechinyere Njoku (Member, IEEE) received the master's degree in aeronautics, mechanical and electronics engineering from Kumoh National Institute of Technology, Gumi, South Korea, in 2021.

She is a Doctoral researcher with the Networked Systems Laboratory, Kumoh National Institute of Technology. Her research interests are in data-driven intelligent transportation systems, digital twin, intelligent energy management systems, and metaverse for the industry.

Ms. Njoku serves as a Reviewer for the *IET Communications Journal*. She is a member of IEEE Young Professionals, and an IEEE Women In Engineering. She is also a member of the Society of Petroleum Engineers, WomenTech Network, and Nigerian Society of Engineers Nigeria.

Cosmas Ifeanyi Nwakanma (Senior Member, IEEE) received the associate degree or national diploma (Distinction) degree in electrical/electronics engineering from the Federal Polytechnic Nekede, Imo, Nigeria, in 1999, the B.Eng. degree in electrical and electronics engineering, the M.Sc. degree in information technology, and the M.B.A. degree in project management technology from the Federal University of Technology, Owerri, Nigeria, in 2004, 2012, and 2016, respectively, and the Ph.D. degree in IT-convergence engineering from the Networked System Laboratory, Department of IT-Convergence Engineering, Kumoh National Institute of Technology, Gumi, South Korea, in 2022.

He has been a Postdoctoral Fellow with the Smart Grid Resiliency and Analytics Laboratory, the Lane Department of Computer Science and Electrical Engineering, West Virginia University, Morgantown, WV, USA, since November 2024. He was an Intern with Asea Brown Boveri, Ilupeju, Nigeria, in 2003. From 2009 to 2019, he was a Lecturer and a Researcher with the Federal University of Technology. From 2022 to September 2024, he was a Senior Research Fellow with the ICT Convergence Research Center, Kumoh National Institute of Technology. His research interests include the intersection of explainable artificial intelligence, Internet of Things, digital twin, and cybersecurity of smart systems, such as the smart grids, factories, homes, farms, and vehicles.

Dr. Nwakanma is a member of the Computer Professionals Registration Council of Nigeria, Nigeria Society of Engineers, and registered by the Council for the Regulation of Engineering in Nigeria.

Jae Min Lee (Member, IEEE) received the Ph.D. degree in electrical and computer engineering from Seoul National University, Seoul, South Korea, in 2005.

From 2005 to 2014, he was a Senior Engineer with Samsung Electronics, Suwon, South Korea, where he was a Principal Engineer from 2015 to 2016. Since 2017, he has been an Associate Professor with the School of Electronic Engineering and the Department of IT Convergence Engineering, Kumoh National Institute of Technology, Gumi, Gyeongbuk, South Korea. Since 2024, he has been the Director of the Smart Defense Logistics Innovation Convergence Research Center. His current main research interests are smart IoT convergence application, industrial wireless control network, UAV, metaverse, and blockchain.

Dr. Lee is the Executive Director of the Korean Institute of Communications and Information Sciences.

Dong-Seong Kim (Senior Member, IEEE) received the Ph.D. degree in electrical and computer engineering from Seoul National University, Seoul, South Korea, in 2003.

From 1994 to 2003, he was a Full-Time Researcher with the ERC-ACI, Seoul National University. From March 2003 to February 2005, he served as a Postdoctoral Researcher with the Wireless Network Laboratory in the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY, USA. From 2007 to 2009, he was a Visiting Professor with the Department of Computer Science, University of California at Davis, Davis, CA, USA. He is currently a Professor with the Department of IT Convergence Engineering at the School of Electronic Engineering, Kumoh National Institute of Technology, Gumi, South Korea. He is also the Director of the KIT Convergence Research Institute and the ICT Convergence Research Center (ITRC and NRF Advanced Research Center Program), supported by the Korean Government, Kumoh National Institute of Technology, and the Director of NSLab Company Ltd., Gumi. His primary research interests include realtime IoT and smart platforms, industrial wireless control networks, networked embedded systems, fieldbus, metaverse, and blockchain.

Prof. Kim served as the Dean for IACF from 2019 to 2022. He is a Senior Member of ACM.