ITU Journal of Wireless Communications and Cybersecurity

ITU/JWCC

Explainable and Trustworthy Artificial Intelligence in 6G THz Networks: Challenges, Solutions, and Future Perspectives

Amine Gonca Toprak[1], **Öykü Berfin Mercan**[1], **Yasin Emre Tok**[1] and **Sümeye Nur Karahan**[1]
[1] R&D Department, TT Mobile Communication Services Inc., Ankara, 06103, Turkey

Abstract: 6G Terahertz (THz) communication technologies are emerging as one of the fundamental cornerstones of next-generation wireless networks, offering high data transmission speeds, ultra-low latency, and dense connectivity. Artificial intelligence (AI) plays a critical role in managing these networks in a high-performance, dynamic, and flexible manner. However, AI models pose significant challenges, especially regarding security, transparency and reliability, due to their opaque decision-making processes. This review provides a comprehensive examination of Explainable Artificial Intelligence (XAI) and Trustworthy AI approaches tailored for 6G THz networks. The study begins by evaluating the role of AI in key use cases such as beamforming, resource allocation, and channel modeling. Then it explores popular XAI techniques, including SHAP(SHapley Additive exPlanations), LIME (Local Interpretable Model-Agnostic Explanations), and attention-based visualization, and discusses their applicability within complex network architectures. Moreover, it investigates critical security threats, such as adversarial attacks, data poisoning, and privacy breaches, and reviews existing solutions aimed at enhancing model robustness and accountability. Finally, this work identifies key challenges associated with the performance-explainability trade-off and outlines promising directions for future research in the development of secure, transparent, and regulation-compliant AI systems for 6G networks.

Keywords: 6G THz networks, explainable artificial intelligence (XAI), trustworthy artificial intelligence, next generation wireless networks.

6G THz Ağlarında Açıklanabilir ve Güvenilir Yapay Zeka: Zorluklar, Çözüm Yaklaşımları ve Gelecek Perspektifleri

Özet: 6G Terahertz (THz) haberleşme teknolojileri, yüksek veri iletim hızları, ultra düşük gecikme süresi ve yoğun bağlantı kapasitesi sunarak gelecek nesil kablosuz iletişim ağlarının temel yapı taşlarından biri olarak öne çıkmaktadır. Bu ağların yüksek performanslı, dinamik ve esnek bir şekilde yönetilmesinde yapay zekâ (YZ) hayati bir rol üstlenmektedir. Ancak YZ modelleri, karar alma süreçlerinin opak yapısı nedeniyle özellikle güvenlik, şeffaflık ve güvenilirlik açısından önemli zorluklar barındırmaktadır. Bu derleme çalışması, 6G THz ağları için özelleştirilmiş Açıklanabilir Yapay Zekâ (XAI) ve Güvenilir Yapay Zekâ (Trustworthy AI) yaklaşımlarını kapsamlı bir şekilde incelemektedir. Çalışma, hüzmeleme, kaynak tahsisi ve kanal modelleme gibi temel kullanım senaryolarında YZ'nin rolünü değerlendirerek başlamaktadır. Ardından, SHAP (SHapley Additive exPlanations), LIME (Local Interpretable Model-Agnostic Explanations) ve dikkat (attention) tabanlı görselleştirme gibi yaygın XAI teknikleri ele alınmakta ve bu yöntemlerin karmaşık ağ mimarilerinde uygulanabilirliği tartışılmaktadır. Bununla birlikte, düşman saldırılar (adversarial attacks), veri zehirleme (data poisoning) ve mahremiyet ihlalleri gibi kritik güvenlik tehditleri incelenmekte; model dayanıklılığı ve hesap verebilirliğini artırmaya yönelik mevcut çözüm önerileri değerlendirilmektedir. Son olarak, performans–açıklanabilirlik dengesi ile ilişkili temel zorluklar tanımlanmakta ve 6G ağlarında güvenli, şeffaf ve regülasyonlarla uyumlu YZ sistemlerinin geliştirilmesine yönelik gelecek araştırma alanları ortaya konulmaktadır.

Anahtar Kelimeler: 6G THz ağlar, açıklanabilir yapay zeka (XAI), güvenilir yapay zeka, yeni nesil kablosuz ağlar.

REVIEW PAPER

Corresponding Author: Amine Gonca Toprak, aminegonca.toprak@turktelekom.com.tr

Reference: A. G. Toprak, Ö. B. Mercan, Y. E. Tok, and S. N. Karahan, (2025), "Explainable and Trustworthy Artificial Intelligence in 6G THz Networks: Challenges, Solutions, and Future Perspectives," *ITU-Journ. Wireless Comm. Cyber.*, 2(2), 61–80.
Submission Date: Apr, 25, 2025
Acceptance Date: Sep, 26, 2025
Online Publishing: Sep, 30, 2025

1 INTRODUCTION

Wireless communication technologies have evolved from 1G analog systems to 5G and beyond. Initially, these systems enabled voice communication; over time, with the increasing data transmission capacity, they paved the way for mobile internet, video streaming, the Internet of Things (IoT), and smart cities [1]. With 4G, broadband internet access has become widespread, and 5G has become usable in critical areas such as industrial automation, autonomous vehicles, and remote surgery with its low latency, high speed, and network slicing. However, despite these developments, wireless communication systems face various challenges. Increasing device density and data traffic lead to inadequate spectrum resources; heterogeneous network structures, security vulnerability, low energy efficiency, and lack of transparency threaten the sustainability of the systems. In this context, it is of great importance that criteria such as not only high performance but also security, explainability, and reliability are met in next-generation communication systems [2], [3].

6G technology developed to overcome these challenges and capacity limitations in wireless communication aims to provide ultra-high transmission speed, latency under milliseconds, and large-scale connection density with the usage of higher frequency bands such as THz spectrum [4]. 6G does not consist of faster communication only, but multidimensional features such as sensing, positioning, and AI-supported network management [5]. In this new architecture, AI systems enable complex processes to be managed autonomously, such as network adaptation to dynamic circumstances, resource allocation, channel modeling, and beamforming (Figure 1). However, this integration brings with it new challenges as well as new opportunities. The unexplainable decision structures of AI models lead to serious concerns about security vulnerabilities and system reliability, which becomes especially critical in sensitive frequency environments such as THz. In this context, the need for explainable and reliable AI approaches is increasing for the secure, transparent, and ethical operation of 6G networks [6].

The ultra-high speeds, low latency, and dense connectivity offered by 6G THz networks necessitate the integration of AI-driven autonomous decision-making mechanisms. However, the “black box” nature of deep learning models employed in these systems introduces significant risks, particularly in critical areas such as security, privacy, and transparency. Ensuring the explainability and traceability of decisions related to the management of critical infrastructures, such as 6G networks, is essential for fostering user trust and meeting regulatory compliance requirements. Therefore, systematically exploring explainable artificial intelligence (XAI) approaches that balance both performance and reliability within 6G THz networks addresses a pressing gap in the current literature and is vital for the safe and trustworthy deployment of next-generation communication systems.

THz communication technology is one of the innovative components to be used in the physical layer of 6G [7]. This frequency band, covering the range of approximately 0.1–10 THz, offers a much wider spectrum compared to the millimeter wave (mmWave) frequencies used in existing cellular systems and has the potential for data transmission at terabit/second levels [8]. Thus, intensive applications such as holographic communication, augmented reality (AR), and real-time high-resolution video streaming can be supported. However, THz wave propagation is seriously affected by atmospheric absorption, free space attenuation, and sensitivity to objects. Therefore, THz communication has some limitations such as short range, high directivity and line of sight requirement. These limitations necessitate advanced engineering solutions such as dynamic beamforming, intelligent reflective surfaces (IRS), directional antenna arrays, and complex channel modeling methods [9]. In such an environment, efficient and stable operation of the network requires the implementation of agile and learning systems rather than traditional deterministic methods.

Although AI approaches offer significant potential for autonomous management of complex processes in 6G THz networks, the decision-making processes of these systems are not transparent, leading to significant security and reliability issues [10]. Strong AI models such as deep learning (DL) achieve high performances on beamforming, resource allocation, and channel prediction tasks [11], however, they lack explanation of how the decision is made. Especially high range, unexplained resolutions, such as THz may lead to risks such as service quality deterioration, expected network degradation, and security vulnerabilities. In addition, AI systems are also vulnerable to threats such as adversarial attacks, model poisoning, data leakage, and biased conflicts [12]. Therefore, not only accuracy, but also transparency, security, and ethical principles-based performance should be part of the parts of AI-based network management. These flexible, XAI and trustworthy AI solutions have become a critical need for 6G THz results to be operated in a secure, traceable and sustainable manner.

XAI is an approach that aims to make the decision-making processes of AI models understandable, interpretable, and controllable by humans [13]. It is challenging to understand why such systems make specific decisions due to the black-box nature of DL-based models. This threatens principles that are critical to 6G THz networks, such as security, transparency, and accountability. XAI techniques aim to overcome this difficulty by increasing the understandability of model outputs and providing insights into system behavior by visualizing the effects of input features on decisions. Among the frequently used methods in the literature are techniques such as SHAP (SHapley Additive exPlanations), LIME (Local Interpretable Model-

Agnostic Explanations), and attention visualizations. With these methods, it is possible to analyze how much weight the model assigns to each input variable and to identify the source of possible erroneous decisions [14]. However, the computational costs of these techniques limit their applicability, especially in resource-constrained edge devices. This makes balancing explainability and performance in dynamic and heterogeneous environments such as 6G systems a challenging engineering problem [15].

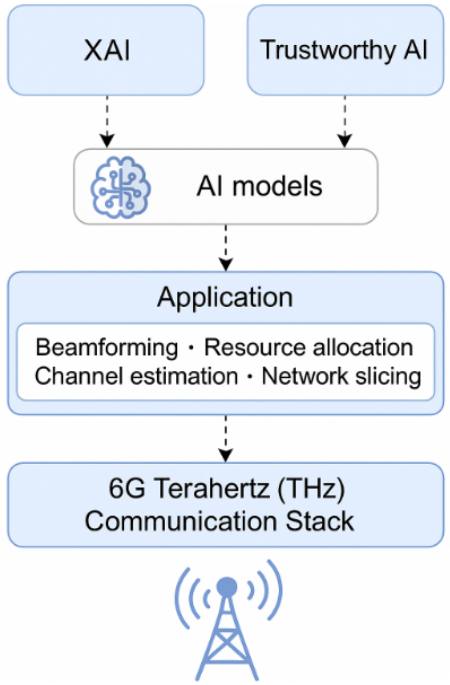

Fig. 1 The integration of Explainable AI (XAI) and Trustworthy AI within the 6G protocol stack is shown. XAI improves transparency, while Trustworthy AI ensures reliability and security for network operations.

XAI techniques make significant contributions by increasing the understandability of AI systems; however, this understandability is not sufficient in environments like 6G THz networks. It is necessary not only to understand how the model works but also to ensure that it works in a way that is safe, fair, robust, and privacy-preserving. At this point, the concept of Trustworthy AI becomes essential, aiming to make systems not only transparent but also secure, ethically aligned, and resistant to adversarial threats [16]. In particular, the widespread use of federated learning, edge AI, and autonomous decision support mechanisms in 6G networks makes threats such as adversarial attacks, model poisoning, data leakage, and biased decision-making more apparent. To provide integrated solutions to these threats at both technical and ethical levels, it is of great importance that XAI approaches are designed in alignment with Trustworthy AI principles [17]. However, this holistic structure also necessitates establishing difficult balances between explainability and performance, and security and flexibility. Thus, this issue becomes not only a technical need but also a multidisciplinary engineering and ethical challenge.

Recent studies reveal that the concepts of explainability and reliability play a critical role not only in theoretical approaches but also in practical system designs. For instance, [18] proposed a security framework supported by XAI techniques to solve the security challenges encountered in IoT systems integrated with 6G networks. The authors applied tree-based machine learning algorithms such as Random Forest, XGBoost and KNN to classify IoT network traffic using the CIC-IoT-2023 dataset. The SMOTE technique was used to eliminate data imbalance then the understandability of model decisions was increased with SHAP and LIME explainability methods. Model insights were cross-validated by comparing feature importance scores with XAI outputs, thus testing the consistency of the model. The obtained results show that tree-based models supported by XAI provide both high accuracy and transparency, providing a reliable solution for 6G-based IoT security systems.

Similarly, studies focusing on trustworthy and XAI applications in 6G-supported edge-cloud ecosystems are drawing attention. In this context, [19] proposed a five-layer architecture that will simultaneously provide explainability and reliability features of AI systems running on edge computing infrastructure. In the study, AI models operating at the edge and 6G networks are defined and the basic security, privacy, hardware and interoperability issues in this structure are detailed. The authors presented a five-level structure in the architecture designed with the XAI-as-a-Service approach: infrastructure, routing, analysis, control and application layers. Within the scope of this structure, XAI components integrated with classical AI operations produce explanations at both local and global levels and the human-centered understanding structures these explanations. The study suggested that the proposed architecture will contribute to the widespread and secure adoption of 6G by increasing trust and transparency, especially in critical applications such as healthcare, autonomous transportation and industrial automation.

In addition, studies developed practical frameworks for evaluating the reliability of DL models in 6G-based applications also stand out in the literature. In this context, [20] presented an approach that evaluates the reliability of two Deep Neural Network (DNN) models in the automatic modulation recognition task, which is an important problem in 6G technology. Classification was performed on a synthetic dataset belonging to the THz band using CNN and ResNet architectures, then analysis was performed on three basic metrics, namely data robustness, parameter sensitivity and resistance to adversarial attacks, within the framework of the developed reliability model. The ResNet model ex-

hibited higher accuracy and stronger performance against noise compared to CNN. In addition, the weak points of the models were determined in the evaluations made with single-bit corruption scenarios for weight parameters and eight different attack types (e.g., PGD, DeepFool, C&W). The study presents one of the first application-oriented reliability analysis frameworks by revealing the necessity of testing DNN-based classifiers used in the 6G environment not only for accuracy but also for reliability.

This study [21], drew attention to the transparency problems in the decision-making processes of AIe applications in 6G wireless communication systems and presented a comprehensive review evaluating XAI-based solutions. DL models that achieve a high accuracy rate, especially on PHY and MAC layers may cause trust weakness due to unexplainable decision structures are emphasized. In this context, a multi-layer explainability framework is proposed using visual explanations, hypothesis tests and local modeling techniques. Following these theoretical approaches, studies showing how explainability and reliability principles are applied in field-oriented systems also attract attention in the literature.

All these developments reveal that 6G and THz communication systems should go beyond being just high-performance communication infrastructures and work in integration with secure, transparent and ethical AI solutions. The proliferation of AI-based decision support systems in 6G networks has made the concepts of explainability and reliability mandatory at the system level. Studies conducted in this direction show that XAI and Trustworthy AI approaches should be addressed not only at the model level but also within end-to-end architectural designs.

There are several review studies in the literature that address AI and XAI in the context of 6G networks. However, these studies primarily provide broad overviews of AI techniques and only superficially examine the fundamental components of 6G. Moreover, existing research lacks an in-depth analysis of critical issues such as THz band-specific security threats, practical implementation scenarios of XAI techniques, and the trade-off between performance and explainability. This article seeks to address this gap by providing a comprehensive examination of XAI and trustworthy AI approaches for 6G THz networks, focusing on their application domains, security vulnerabilities, proposed solutions, and directions for future research.

This article presents a comprehensive review of the increasing integration of AI in 6G and THz communication systems, focusing on explainability and reliability. Unlike many performance-centric studies in the literature, it provides a holistic perspective on XAI and Trustworthy AI implementations. The main contributions of the study can be summarized as follows:

- The role and application areas of AI in 6G THz networks are investigated in detail, techniques such as reinforcement learning, federated learning and edge AI are evaluated in the context of applications such as beamforming, resource allocation, channel modeling and IRS control.

- The conceptual foundations and application examples of XAI techniques are discussed. How SHAP, LIME and Grad-CAM methods are applied to different AI models, especially through graph-based structures and attention mechanisms, is detailed. The fundamental challenges encountered in providing explainability and making models transparent in reinforcement learning are also addressed.

- From the perspective of Trustworthy AI, vulnerabilities in 6G networks are analyzed. Threats such as adversarial attacks, data poisoning, model leakage and privacy are systematically examined. In addition, advanced security solutions such as blockchain and post-quantum cryptography are discussed.

- The balance problem between performance and explainability is examined, especially in computationally constrained environments such as edge devices. The adaptation difficulties of AI models in dynamic and heterogeneous THz networks are evaluated.

- Solution proposals such as lightweight XAI approaches, white-box/black-box model hybrids, and controllable AI systems are presented. Examples of how these approaches are implemented in industrial application areas are also shared.

- Forecasts on the future of 6G and XAI are presented. XAI-supported 6G standardization studies, ultra-low power consumption explainability frameworks, blockchain-based update systems and XAI integrations in THz scenarios are discussed.

- Finally, open research topics in the literature are identified. Contributions are made to future research directions for the secure, transparent and ethical operation of 6G-THz systems.

The remainder of this paper is structured as follows: Section 2 introduces the application areas of AI in 6G THz networks. Section 3 discusses XAI techniques and their applications. Section 4 examines the concept of trustworthy AI in the context of security, privacy, and attack scenarios. Section 5 discusses the main challenges related to the balance of explainability and performance. Section 6 includes solution approaches and architectural proposals presented in the literature. Section 7 conveys future perspectives and research directions. Finally, Section 8 summarizes the work and presents open research problems.

2 AI APPLICATIONS IN 6G THz NETWORKS

From 1G to 5G, the requirements for communication systems have grown significantly. To address these demands, communication systems have been continuously evolving, and 5G provides numerous vital solutions for applications that demand high data traffic, including high-resolution video transmission, multiplayer online games, live streaming, and augmented and virtual reality (AR/VR). Moreover, with data traffic set to increase even further, by 2030, data traffic per phone is expected to reach over 60 GB per month [22]. Therefore, next-generation communication networks are engineered to address stringent demands for throughput, scalability, latency, and complexity while targeting data rates of up to 1 Tbps [23]. The evolution toward 6G and beyond will be driven by cutting-edge use cases like the AI of things (AIoT), autonomous driving, smart manufacturing, and edge AI. To implement these use cases, new wireless technologies are required [24], and machine learning (ML) tools will be key components [25]. Some of the main areas in wireless communications to which machine learning provides solutions include beamforming, resource allocation, channel estimation, and the optimization of reconfigurable intelligent surfaces (RIS).

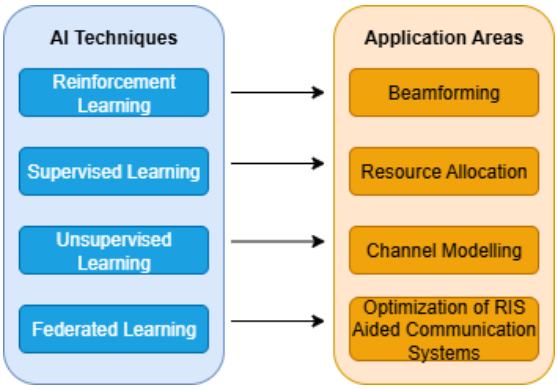

Fig. 2 AI techniques are mapped to different 6G THz network application areas. These methods enable intelligent and adaptive network management for various use cases.

Figure 2 shows the relationship between the AI techniques commonly used in 6G THz communication systems and their application areas. Each AI method contributes to meeting different functional needs of the network architecture. Reinforcement learning is especially effective in tasks such as beamforming, where continuous adaptation to dynamic environments is required. Supervised learning techniques are used to optimize bandwidth, delay, or energy consumption in resource allocation problems using labeled data. Unsupervised learning helps reveal hidden patterns in tasks such as channel modeling without the need for labeling. Finally, federated learning is suitable for scenarios where training is done locally without sharing data, preserving both privacy and compliance, especially in the control of smart RIS. These AI-based solutions play a critical role in enabling the intelligent automation and adaptive flexibility required by 6G systems.

2.1 Machine Learning Based Beamforming

In mmWave and THz MIMO systems, large-scale antenna arrays are deployed to counteract the severe path loss and guarantee adequate received signal power. To address these limitations, a variety of analog, digital, and hybrid beamforming architectures have been developed and rigorously evaluated. Beamforming, a core feature of multi-antenna wireless systems, has delivered notable gains in 5G networks, particularly in terms of energy efficiency and directivity. Despite their widespread adoption, classical beamforming codebooks exhibit three principal limitations. First, achieving full angular coverage requires a large ensemble of narrowly focused beams, which incurs substantial beam-training overhead. Second, the single-lobe patterns optimized for maximum directivity can be suboptimal in non-line-of-sight (NLOS) scenarios, where more diffuse beam shapes may yield better coverage. Third, these codebooks generally presuppose a fully calibrated array with known geometry-an assumption that not only entails high calibration costs but also impedes deployment on platforms with arbitrary or imperfectly characterized antenna layouts [26]. These limitations originate from the systems' lack of adaptability to the propagation environment and inherent hardware constraints. To overcome these challenges, deep learning and deep reinforcement learning models can be employed. Thanks to recent advances in AI, machine-learning–based beamforming has attracted considerable attention for its ability to mitigate the high overhead and complexity inherent in classical designs. In [26], a deep reinforcement learning based approach was developed to adapt the codebook beams to the environment. In [27], the authors trained a convolutional neural network on the sub-6 GHz band to predict the optimal mmWave beam, significantly reducing beam-training overhead. In [28], the authors proposed a long short-term memory (LSTM) recurrent neural network to predict transmit beamforming vectors from historical channel data, thereby enhancing channel estimation accuracy.

2.2 Channel Modeling Using Machine Learning

In 6G THz networks, AI-driven channel modeling has emerged to overcome the limitations of traditional measurement- and statistics-based approaches. Generative adversarial networks (GANs) have been applied to synthesize realistic THz channel parameters: e.g., transfer-learning-enabled, transformer-based GAN (TT-GAN) models can generate high-fidelity channel realizations from limited measurement data, achieving precise power-delay pro-

files with low RMSE and high structural similarity, while transfer-GAN (T-GAN) frameworks fine-tune pre-trained models on small measurement sets to closely match empirical channel distributions [29].

2.3 Resource Allocation Using Machine Learning

Continued advancement of 5G and future-generation networks will allow both the network architecture and the underlying communication technologies to be reconfigured on the fly to suit evolving demands. However, achieving this level of adaptability incurs substantial signaling overhead and computational burden to coordinate resources efficiently. Therefore, ML-based approaches could be helpful to redefine the resource allocation problem. In 6G THz networks, resource allocation has been successfully reformulated as a machine-learning problem in which agents learn to assign spectrum and power in real time. Deep reinforcement learning (DRL) techniques cast the joint subcarrier-and-power assignment task as a Markov decision process, with neural-network function approximators—such as the soft actor-critic (SAC) model—efficiently balancing throughput, signaling overhead, and computational complexity by dynamically switching between centralized and distributed allocation modes in smart soft-RAN architectures. Deep reinforcement learning–based approaches provide a throughput–overhead–complexity (TOC) advantage over centralized or distributed fixed schemes [30]. More broadly, various DRL frameworks allow agents to interact with stochastic THz environments and refine allocation policies via trial-and-error feedback, achieving near-optimal performance without explicit channel models. To address privacy and scalability in ultra-dense deployments, federated learning enables geographically distributed edge nodes to collaboratively train shared resource-allocation models without exchanging raw channel measurements, thereby reducing signaling overhead while preserving user data confidentiality [31].

2.4 Optimization of RIS Aided Communication Systems Using AI

RIS are among the key enabling technologies for beyond-5G (B5G) and 6G networks. Composed of a large array of passive elements, an RIS can dynamically shape the radio environment by reflecting incident signals toward desired directions [32]. However, RIS-assisted systems introduce new challenges compared to conventional architectures—namely, acquiring accurate channel state information (CSI), optimizing phase shifts, and determining optimal RIS placement. To address these issues, a variety of DL and DRL techniques have been proposed [33], [34]. DL techniques—e.g., convolutional neural networks (CNNs), deep denoising CNNs (DnCNN), and multi-layer perceptrons—have been used for high-fidelity cascade channel estimation and direct mapping from received pilots to optimal phase-shift and beamforming configurations, achieving near-optimal spectral efficiency with dramatically reduced pilot overhead. Moreover, federated learning frameworks distribute model training across IRS-equipped edge nodes to preserve user privacy and reduce signaling overhead, while still converging to global beam-reflection and resource-allocation policies that match centralized performance [35].

3 CONCEPTUAL AND PRACTICAL ASPECTS OF XAI TECHNIQUES

AI applications have become increasingly prevalent in next-generation communication technologies, including THz communication, supporting tasks such as beamforming [36], [37], [38], channel estimation [39], [40], [41], and resource allocation [42], [43], [44]. However, the decision-making processes of these AI-driven systems often function as black boxes, making it difficult to understand how specific decisions are made. At this point, XAI comes into play, offering transparency and reliability for AI-based next-generation communication technologies [15], [45]. XAI is a field of study that aims to explain the decisions made by AI systems in a way that is understandable to humans [46]. Despite the remarkable performance of AI systems, they often operate as black boxes, and it is not clear how these models reach decisions. XAI offers a transparent and reliable framework that ensures AI systems are explainable and trustworthy. In the domain of next-generation wireless communication, XAI aims to enhance system reliability, security, and service quality, strengthen ethical compliance, and increase trust in autonomous operations by making the complex structure of AI-based models more transparent [12], [15]. XAI plays a key role in enhancing various tasks in next-generation communication, such as beamforming, resource allocation, and signal modulation [15], [16], [47].

3.1 Explainable AI Techniques: Interpreting Complex Models with SHAP, LIME, and Grad-CAM

To address interpretability issues of complex AI models, various XAI methods have been proposed in Figure 3. Although, these methods differ in terms of their developed approaches, the aim is the same make the decision processes of the models explainable. The literature presents several core XAI techniques that are developed specifically for different model types, data structures and usage scenarios that try to make the decisions of the AI system transparent. One of the fundamental techniques is SHAP [48], which calculates the individual contribution of each feature to the model's decision based on Shapley values from cooperative game theory. This method provides model-agnostic, fair, and theoretically grounded explanations. Another prominent technique is LIME [49], which

approximates a complex model's local behavior around a specific prediction using a simpler surrogate model.

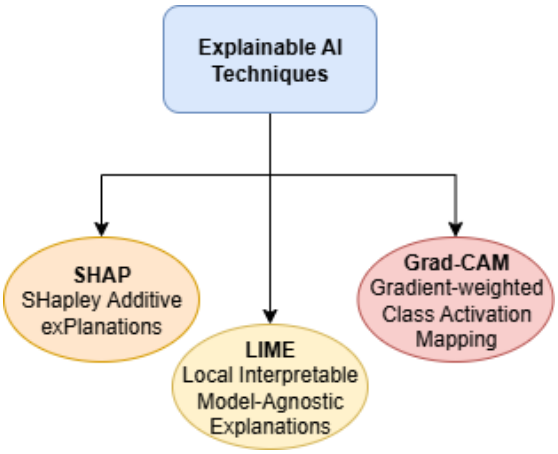

Fig. 3 Key Explainable AI (XAI) techniques are illustrated, including SHAP, LIME, and Grad-CAM. These methods help interpret and visualize complex model decisions.

LIME generates a local dataset by perturbing the input around the target prediction. A simple interpretable model is then trained on this dataset to mimic the behavior of the complex model within that local region. Another widely used method is Grad-CAM (Gradient-weighted Class Activation Mapping) [50], which provides highly effective explanations for image classification. It visualizes the regions of the image that deep learning approaches, such as CNN focus on while making image class predictions with a heat map. This method is a valuable method for error analysis, allowing you to observe the areas where the model misunderstood. XAI methods that focus on different tasks are integrated into next-generation wireless communication systems to explain the decision-making processes of AI models addressing various tasks and to enhance the transparency and interpretability of system behavior. For instance, SHAP and LIME methods of XAI were proposed to detect cyber threats in complex 6G environments [6], while another study [18] proposed to validate model feature prediction.

Another study employed Grad-CAM to visualize and explain beam direction predictions in a spatial attention-based model [51]. In systems using transformer-based architectures, attention-based visualization methods allow identification of which input features the model focuses on during inference [52]. These visualizations help analyze the underlying structure of the decision-making process by showing how attention is distributed across inputs.

3.2 Explainability Techniques for Graph Neural Networks

The need for explainability extends to Graph Neural Networks (GNNs), which are increasingly used in AI-driven wireless communication systems [53]. GNNs are deep learning models that operate on graph-structured data, capturing the relationships between nodes. To interpret their complex architectures, several XAI methods have been proposed, including attention-based, subgraph-based, edge-based, gradient analysis-based, and feature-based techniques.

Among these, attention-based methods stand out by offering insight into how the model processes node relationships. Attention mechanisms in GNNs assign importance to neighboring nodes, leading to more interpretable results [54].

Beyond attention methods, various explainability approaches have been proposed: GNN Explainer [55] identifies the subgraphs, edges, and node features most relevant to a model's prediction. SubgraphX [56] performs probabilistic extraction of influential subgraphs. PGExplainer [57] focuses on learning edge importance through a probabilistic framework. Gradient-based methods, such as saliency maps and integrated gradients, analyze the influence of inputs based on backpropagated gradients [58], [59]. GraphLIME [60] applies a linear surrogate model to explain individual node features. RelEx [61] explains local feature importance in the learned representations. In addition, SHAP has been adapted for GNNs in the form of GraphSHAP, analyzing the contributions of nodes and subgraphs using Shapley values [62].

GNN-based methods have been proposed in numerous studies focused on 6G and THz communications [63], [64], [65], [66], [67]. As the demand for reliable and transparent AI grows, XAI remains an open research area in GNNs. For example, [68] proposed Graph Reinforcement Learning with an emphasis on explainability, and introduced Bayesian GNN Explainer, a method for analyzing edge and node influence in model decisions.

3.3 Explainability Techniques in Reinforcement Learning

In next-generation communication technologies, RL methods are introduced due to their adaptability to complex and dynamic decision-making processes. RL offers solutions for tasks such as routing, resource allocation, and channel estimation, outperforming traditional methods in many cases [69], [70], [71], [72], [73]. However, the inherent characteristics of THz systems make it challenging to interpret the decisions made by RL agents. At this point, XAI techniques play a critical role in making these systems more interpretable and trustworthy. RL is an AI approach that enables agents to learn through interactions with their en-

vironment by using reward signals [74]. This learning process often remains a black box, meaning it is not easy to understand why agents make certain decisions or adopt specific strategies. XAI aims to improve transparency, auditability, and reliability in the decision-making processes of RL systems [75]. Although traditional XAI methods such as SHAP and LIME are sometimes applied in RL systems, their direct applicability is limited [15]. These methods are effective in supervised learning scenarios, but due to the temporal dynamics of RL—where actions are influenced by both past experiences and future rewards—they may not always be suitable. Thus, explainability in RL remains a major area of ongoing research.

Various specialized XAI approaches have been proposed for RL in the literature: Programmatically Interpretable RL (PIRL) [76] learns agent policies in a human-interpretable way using policy sketches, providing a constrained and explainable alternative to black-box deep RL policies. Hierarchical Policies [77] decompose complex tasks into simpler sub-tasks. These are either solved by reusing previously learned policies or by learning new ones as needed. Linear Model U-Trees (LMUT) [78] is a post-hoc explanation method that uses linear models at the leaf nodes of a decision tree to simulate Q-functions, making the behavior of complex RL systems interpretable. These techniques offer different levels of transparency and serve to bridge the gap between the performance of RL agents and the need for trustworthy AI in 6G systems.

3.4 Challenges and Risks in Achieving Explainability

Although AI models can deliver high accuracy, their complex and opaque structures make them difficult to interpret. Transparency and explainability are essential for trustworthy systems, particularly in high-risk applications like 6G communication networks [15]. However, several technical and structural challenges arise when increasing transparency in such complex and rapidly evolving systems. One major challenge is the lack of standardization, which hinders comparability, reproducibility, and reliability across different applications. Many developed XAI techniques are tailored to specific use cases, making it difficult to generalize their applicability to broader systems like 6G networks [12], [79].

Another key issue is the accuracy–speed trade-off [15]. In real-time 6G applications, models must produce both accurate and explainable results with minimal latency. However, the computational overhead of complex XAI methods—such as SHAP and LIME—introduces significant delays, posing challenges for their integration into time-sensitive systems [12]. While edge computing in 6G brings advantages like distributed processing, it also amplifies computational constraints. Resource-limited edge devices struggle to handle the intensive calculations required by traditional XAI techniques. On the other hand, faster but simpler XAI methods often produce shallow or less informative explanations, which may not meet the needs of critical applications.

The task-model compatibility issue is another concern. In complex tasks, applying XAI methods such as SHAP and LIME may introduce latency due to computational demands. Moreover, these methods are primarily designed for static models, limiting their effectiveness in RL-based dynamic systems [15]. A critical concern in deploying explainable models is balancing explainability and privacy. Exposing internal model logic may increase the system's attack surface, leading to potential data leakage and security threats. In the context of 6G, where privacy is paramount, ensuring that XAI methods are secure and ethically aligned is essential [80].

In summary, explainability in AI is not just a technical challenge but a multidimensional issue involving ethical, legal, and security considerations. Designing XAI methods that are efficient, privacy-preserving, and suitable for dynamic environments like 6G networks remains an open and urgent research problem.

4 SECURITY, PRIVACY AND RELIABILITY IN TRUSTWORTHY AI

As 6G THz networks increasingly rely on AI for critical functions—including rapid beamforming, resource allocation, dynamic channel modeling, and RIS control—the overall performance and dependability of these systems hinge on the trustworthiness of their AI components. Ensuring secure, private, and reliable AI operation is therefore indispensable. Hostile actors may exploit vulnerabilities at both the training and inference phases, while distributed architectures must also guard against insider threats and prepare for adversaries in the quantum era. This section highlights four key dimensions of trustworthy AI in this context:

- **Adversarial Attacks and Defensive Mechanisms:** DNNs deployed for beamforming and channel prediction in mmWave/THz systems are highly susceptible to adversarial perturbations, where even imperceptible changes in input can cause significant misalignment and throughput degradation [81]. To mitigate these risks, adversarial training—exposing models to crafted adversarial examples during training—and defensive distillation—which smooths the network's output distributions—have been shown to improve robustness against both fast-gradient and iterative attacks [82].

- **Data Poisoning and Model Leakage Threats:** Distributed training workflows in wireless networks are vulnerable to data poisoning and backdoor attacks, where malicious inputs or updates can corrupt the global model in federated learning settings [83]. In addition, membership inference attacks exploit gradient

or output observations to determine whether a specific user or channel sample was part of the training dataset, thereby violating data confidentiality [84].

- **Privacy-Preserving AI via Federated and Differential Privacy:**
  To protect raw THz channel data and user-specific information, federated learning enables model training at edge nodes without sharing raw data, transmitting only aggregated updates [85]. In parallel, differential privacy introduces mathematically calibrated noise into gradients or model parameters, offering formal guarantees on information leakage while preserving utility [86].

- **Blockchain and Post-Quantum Cryptography for Model Integrity:**
  Blockchain technology can be employed to create tamper-proof ledgers that record model update hashes and access control policies, deterring unauthorized modifications in decentralized AI systems [5]. Additionally, post-quantum cryptographic methods—including lattice-based encryption and hash-based signature schemes—can secure AI model distribution and control-plane communications against emerging threats from quantum-capable adversaries [87].

5 CHALLENGES OF EXPLAINABILITY-PERFORMANCE TRADE-OFF

With the rapid advancements in AI, the adoption of AI-based solutions across various domains has accelerated, leading to the development of highly accurate but increasingly complex models whose decision-making processes are often difficult to interpret. In parallel, research into model transparency and explainability has also gained significant momentum. As a result, the explainability–performance trade-off has emerged as a critical issue in XAI studies [88], [89], [90], [91], [92].

5.1 Balancing Accuracy and Transparency: Post-hoc and Intrinsically Interpretable Models

The explainability–performance trade-off presents a major challenge, particularly in applications where model transparency and interpretability are essential. There is an inherent balance between the complexity of AI models and their interpretability. Models such as DNNs, ensemble systems, and attention-based architectures can achieve high accuracy and generalization performance, yet they tend to function as black boxes [88]. Consequently, the lack of interpretability in these models can lead to inconsistencies and trust issues in high-stakes applications [89].

Two principal modeling approaches have emerged to address this trade-off: post-hoc explainability methods and intrinsically interpretable models.

- Post-hoc methods attempt to explain the decisions of already trained complex models by analyzing their input–output behavior, without accessing or modifying their internal mechanisms [93]. Techniques such as LIME, SHAP, and Grad-CAM fall into this category. However, since these approaches are approximative, they may not fully capture the original model's logic, which can compromise reliability in critical applications [89].

- In contrast, intrinsically interpretable models offer transparency within their structure. Their decision mechanisms are directly observable and analyzable. Examples include decision trees, linear regression, and logistic regression models [94]. Although these models typically yield lower performance than complex architectures, they provide faster and more trustworthy insights—especially valuable in safety-critical domains.

5.2 Adaptation Challenges for AI Models in Dynamic and Heterogeneous THz Networks

THz communication systems operate at frequencies beyond millimeter waves, offering distinct advantages such as ultra-high data rates, low latency, and wide bandwidth [95]. These features make THz systems highly promising for 6G and beyond. However, they are also characterized by a dynamic and heterogeneous structure with inherent physical limitations—including high path loss, atmospheric and material absorption, sensitivity to blockages, and strong directionality [24], [96], [97], [98].

Channel conditions in THz environments fluctuate rapidly, influenced by environmental dynamics, user mobility, steerable antenna configurations, and multiuser access scenarios [99]. This high degree of variability necessitates the development of agile, adaptive, and high-performance AI models capable of operating reliably under diverse and shifting conditions. In response, recent literature has explored approaches such as DL, domain adaptation, and transfer learning to adapt AI models to such heterogeneous environments [39], [100], [101], [102]. Accordingly, the development of flexible, low-latency, and XAI solutions has become critical for THz networks.

The adaptation of AI models to real-time channel conditions in THz systems makes understanding decision-making processes more complex. In this context, XAI techniques provide transparency by explaining which channel characteristics the model is affected by and enable human intervention [45]. Therefore, the integration of XAI into these systems is critical not only for performance improvement but also for reliability and traceability.

6 APPROACHES TO ADDRESS KEY CHALLENGES

In this section, we examine solution approaches in the current literature that aim to address the explainability and reliability challenges of AI systems deployed in 6G and THz networks. In particular, we focus on hybrid architectures that combine white-box and black-box models, lightweight XAI techniques suitable for resource-constrained environments, frameworks for controllable AI systems, and their industrial application scenarios (Figure 4). These approaches are evaluated based on both their theoretical contributions and practical implementations, providing a structured perspective for integrating explainable and reliable AI into future 6G systems.

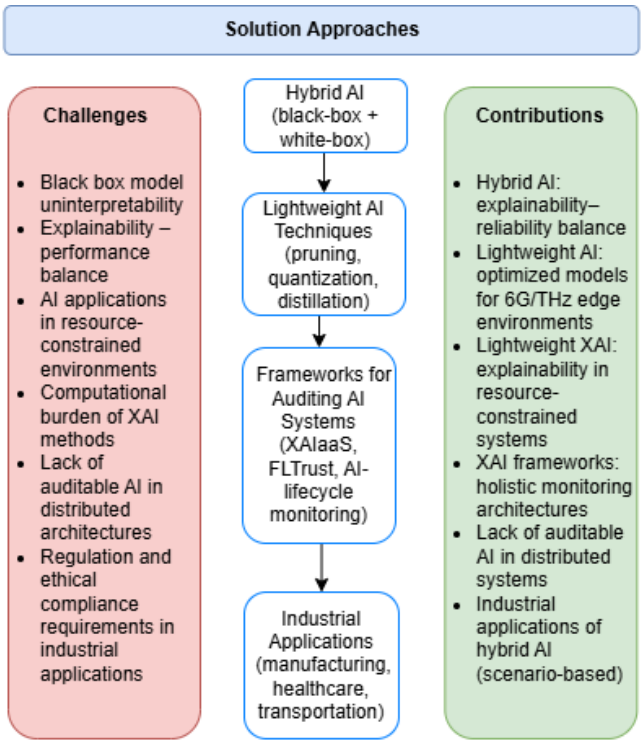

Fig. 4 Challenges, solution approaches, and contributions for XAI in 6G THz networks are summarized. The framework links issues such as transparency, performance, and security to targeted solutions.

6.1 Hybrid AI

Hybrid AI refers to an integrated approach that combines white-box models (interpretable, rule-based) and black-box models (complex but high-performing). White-box models include interpretable structures such as decision trees, linear and logistic regression, and rule-based systems, offering direct insights into their decision-making logic. Conversely, black-box models—such as deep neural networks, support vector machines, and ensemble models like random forests—often provide superior accuracy but lack transparency [103]. The goal of hybrid AI is to combine the strengths of both paradigms, ensuring high performance while maintaining transparency, reliability, and controllability. In this context, the decisions made by a black-box model can be either validated or explained through a white-box model. In some applications, task delegation is used: a white-box model handles low-risk decisions, while the black-box model manages more complex cases.

Hybrid AI is particularly valuable in domains requiring both accuracy and interpretability, such as healthcare, finance, security, autonomous vehicles, and communication infrastructures [104]. These systems can be further enhanced through deep learning models supported by post-hoc explanation tools, attention-based networks, or surrogate model mapping strategies. However, these enhancements may increase system complexity, leading to challenges in ensuring explanation consistency and real-time decision support [105]. Optimizing such hybrid systems thus remains an open research challenge. The integration of explainability and reliability requirements into 6G THz networks introduces various technical and architectural challenges. Literature suggests that effective solutions revolve around maintaining a balance between performance and explainability, ensuring computational efficiency, and developing secure, trustworthy decision-making mechanisms [106].

Recent studies highlight the growing prominence of hybrid AI approaches. For example, [15] emphasizes that XAI techniques should be incorporated directly into system architectures to achieve transparency in AI-based decision mechanisms for 6G. The authors argue that in time-critical, distributed applications—such as network slicing and edge computing—explainability should be meaningful not only for end users but also for system administrators, network operators, and service providers. Post-hoc XAI methods such as SHAP and LIME are proposed to enhance traceability by visualizing which inputs influence model decisions, enabling both real-time and retrospective audits. The authors advocate combining high-performing DNN models with white-box logic to create hybrid structures for decision support systems. Similarly, in [107], hybrid AI systems are analyzed from an Industry 5.0 perspective, discussing human–machine interaction, privacy, and reliable decision-making. The study emphasizes that different levels of explainability should be tailored for different user roles within the Human–Machine Interface (HMI) layer and 6G core network. Furthermore, technologies like federated learning and blockchain are shown to enhance system-level explainability and privacy by offering decentralized, auditable, and transparent architectures. In summary, hybrid AI architectures are emerging as strategic enablers for achieving transparency, accountability, and relia-

bility in heterogeneous, dynamic, and ultra-low-latency 6G application scenarios.

6.2 Lightweight XAI

Lightweight AI refers to optimized AI models that are designed to operate on systems with limited hardware resources. This approach is particularly relevant for 6G and THz network-based edge devices, where ultra-low latency and high-frequency data processing requirements demand compact, efficient models. To achieve this, techniques such as model compression, parameter pruning, quantization, and knowledge distillation are commonly employed to develop models that are smaller, less energy-consuming, and suitable for real-time applications [108].

However, in many critical applications, it is not only the prediction accuracy that matters—understanding why a model makes a given decision is equally important. This is where the concept of Lightweight XAI becomes essential. Lightweight XAI aims to develop computationally efficient explainability methods that can run on edge devices without compromising transparency.

For instance, post-hoc techniques like SHAP and LIME can be adapted into lightweight variants that operate with fewer input samples. In addition, attention visualizations, simplified counterfactual explanations, and feature scoring-based methods aim to offer interpretability while minimizing computational overhead [18]. These techniques allow for system decisions to remain auditable while maintaining energy efficiency and data security. The joint design of lightweight AI and XAI forms a foundation for sustainable, explainable AI solutions tailored for heterogeneous, large-scale, and resource-constrained 6G environments.

However, implementing XAI methods introduces additional computational overhead. While XAI enhances transparency by clarifying the decision-making logic of AI models, it also brings challenges such as increased memory usage, latency, and energy consumption—factors that are particularly problematic for resource-limited edge devices [15]—Commonly used post-hoc methods such as SHAP and LIME demand significant processing power [13], making them impractical for many edge-based deployments. To mitigate these issues, lightweight XAI techniques have been proposed. For example, TreeSHAP [109] is more computationally efficient alternative to standard SHAP. Lightweight XAI frameworks aim to maintain explainability while enabling real-time decision-making in edge environments [110]. Another approach is to use an intrinsically interpretable model that inherently offers explainability and are more suited to low-power edge devices. Although these models may fall short in complex tasks compared to deep networks, they offer critical advantages in scenarios where interpretability and low latency are priorities [94].

A promising direction is the development of hybrid edge–cloud architectures [111], where the AI model operates on the edge device while explainability tasks are offloaded to nearby fog or cloud servers. This division reduces the computational burden on edge devices but introduces new challenges such as data transfer latency, security risks, and scalability concerns [110]. Additionally, selective or on-demand explainability has been introduced. In this approach, explanations are generated only when confidence scores are low, when triggered by user queries, or when predefined thresholds are exceeded. A different strategy is interpretable knowledge distillation [112]. In this approach, a smaller, interpretable "student" model is trained using the outputs of a complex "teacher" model. The student model serves as a transparent surrogate, providing explanations while benefiting from the accuracy of the original.

6.3 Proposed Frameworks for Auditing AI Systems

As AI systems play increasingly critical roles in 6G and THz communication infrastructures, it becomes essential for these systems to not only make accurate decisions but also be controllable, traceable, and reliable. To meet this need, the literature proposes various multi-layered control frameworks that aim to monitor AI performance, interpret decisions, and provide actionable feedback to users at different levels. These frameworks typically consist of three main components:

1. **Model monitoring:** Detects the real-time behavior, deviation and performance degradation of the model.
2. **Explainability layer:** Presents the reasons for model decisions in a form that is understandable to the user or operator.
3. **Threat analysis and security layer:** Detects and responds to risks such as adversarial attacks, model poisoning and data leakage.

In heterogeneous and distributed network architectures like 6G, these frameworks are recommended to be deployed in a decentralized manner, often in combination with federated or privacy-preserving computing in both edge and cloud environments.

For example, the XAI-as-a-Service (XAIaaS) architecture proposed by [15] offers role-specific explanation formats at the user and engineer levels, embedding explainability directly into system infrastructure. Likewise, the FLTrust framework [113] provides a security-focused structure for federated learning, ensuring model update integrity and defending against untrusted client behavior. Other systems, such as AI Lifecycle Monitoring, aim to ensure transparency throughout the model lifecycle, from training to deployment and field use. These frameworks are not only important for model security, but also play a crucial role in ensuring compliance with ethical, legal, and operational regulations [114].

6.4 Industrial Applications

The ultra-low latency, high bandwidth, and massive connectivity capabilities of 6G and THz-based communication systems are enabling advanced AI-driven automation across various industrial domains. In such scenarios, where both decision accuracy and system reliability are critical, hybrid AI approaches have become increasingly prominent. In smart manufacturing, black-box deep learning models are frequently used for real-time quality control, fault detection, and predictive maintenance [115]. However, for tasks requiring human interpretability—such as fault explanations or compliance reports—hybrid AI systems integrating white-box models are preferred.

Similarly, in autonomous driving, high-accuracy perception models that fuse radar and camera inputs must be paired with explainability layers to support safety and accountability [116]. In healthcare technologies, explainable hybrid AI systems are essential for applications such as remote diagnostics, medical imaging, and mobile health, due to strict requirements around patient privacy and decision transparency [79]. The high-resolution and low-latency communication infrastructure provided by the THz spectrum enables these systems to function seamlessly, while hybrid AI ensures that they are also transparent, auditable, and ethically aligned. In critical infrastructure management (e.g., energy, transportation, public security), hybrid AI solutions support early anomaly detection and proactive risk assessment, enabling decisions that are both technically sound and legally traceable.

7 FUTURE RESEARCH DIRECTIONS

In order for 6G THz communication systems to achieve key objectives such as ultra-low latency, high reliability, and dense connectivity, the integration of XAI and Trustworthy AI techniques plays a critical role. This need is particularly pronounced in AI-driven processes such as autonomous network management, channel modeling, beamforming, and resource allocation, where the traceability and justification of AI decisions are becoming indispensable for ensuring system integrity.

Future research is expected to focus on advancing XAI techniques—such as model simplification, feature attribution maps, and local interpretation methods—to improve the interpretability of deep learning-based architectures without compromising performance. Given the computational intensity of many existing XAI methods, a key research priority is the development of ultra-low-power explainability frameworks suitable for edge computing environments.

Additionally, next-generation auditing mechanisms are needed to maintain model transparency and accountability within the distributed and heterogeneous infrastructures of 6G. In this context, blockchain-based structures are gaining traction as promising tools for ensuring secure model updates, verifiability, and protection against adversarial attacks and data integrity breaches (Figure 5). Importantly, future work must address explainability not only at the model level, but also at the system level, aligning with emerging regulatory requirements, ethical standards, and the growing need for user trust in AI-supported 6G applications [117], [118].

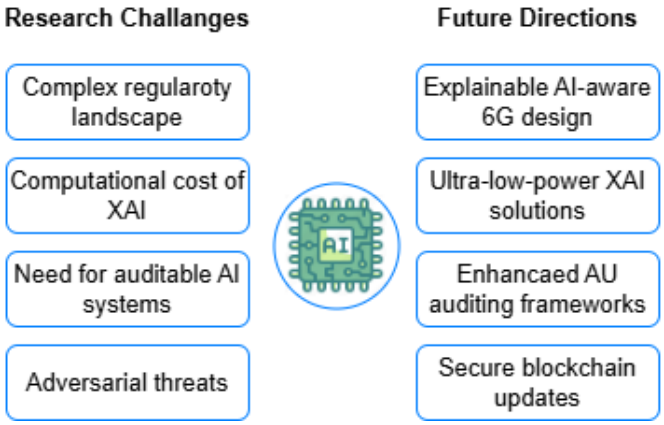

Fig. 5 Open research challenges and future directions for XAI and trustworthy AI in 6G THz networks are presented. These directions highlight emerging areas for secure and explainable network development.

8 CONCLUSION

This study presents a comprehensive literature review on the growing integration of AI in 6G THz communication systems, with a particular focus on the dimensions of explainability and reliability. While the majority of existing research centers on performance optimization, this study offers a holistic perspective, emphasizing the necessity of explainable, trustworthy, and ethically-aligned decision mechanisms in future communication infrastructures. The findings underscore that XAI and Trustworthy AI are not optional, but foundational components for the secure, transparent, and accountable operation of 6G networks. Despite notable progress in this direction, several open research challenges remain. Most notably, the opaque nature of deep neural networks continues to pose serious transparency issues-particularly in safety-critical tasks such as beamforming and channel estimation. General purpose XAI methods often fall short in high-risk service environments, underscoring the need for task-specific and application-aware explainability strategies. Moreover, the trade-off between model performance and interpretability poses significant engineering challenges in resource-constrained edge environments. In distributed learning frameworks, additional concerns such as data leakage, adversarial robustness, and security in federated learning settings further complicate the design of reliable AI systems. In light of

these challenges, the development of advanced AI solutions for 6G requires multidisciplinary collaboration-bringing together communication engineering, machine learning, ethics, and network optimization. Future efforts must focus on designing new methods, tools, and standards that support the creation of AI systems that are not only high-performing, but also secure, explainable, and regulation-compliant.

AUTHOR CONTRIBUTIONS

Amine Gonca Toprak led the study, defined its scope, conducted the literature review, compiled proposed solutions, developed the visual materials, and prepared the manuscript. Öykü Berfin Mercan was responsible for the section on XAI techniques and the "Performance–Explainability Balance" discussion. Yasin Emre Tok prepared the section on AI applications and contributed to the "Trustworthy AI" content. Sümeye Nur Karahan wrote the "Future Perspectives" and "Conclusion" sections and assisted in framing the overall study outputs.

ACKNOWLEDGEMENTS

This work is supported by The Scientific and Technological Research Council of Türkiye (TÜBİTAK) 1515 Frontier R&D Laboratories Support Program for Türk Telekom 6G R&D Lab under project number 5249902.

REFERENCES

[1] A. A. A. Solyman and K. Yahya, "Evolution of wireless communication networks: From 1g to 6g and future perspective," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 12, no. 4, pp. 3943–3950, 2022, ISSN: 2722-2578. DOI: 10.11591/ijece.v12i4.pp3943-3950. [Online]. Available: https://ijece.iaescore.com/index.php/IJECE/article/view/27115.

[2] B. Bakare and E. Bassey, "A comparative study of the evolution of wireless communication technologies from the first generation (1g) to the fourth generation (4g)," *Int J Elect Commun Comp Eng*, vol. 12, no. 3, pp. 73–84, 2021, ISSN: 2249–071X. [Online]. Available: https://www.ijecce.org/index.php/issues?view=publication&task=show&id=1366.

[3] R. Agrawal, "Comparison of different mobile wireless technology (from 0g to 6g)," *ECS Transactions*, vol. 107, no. 1, p. 4799, 2022. DOI: 10.1149/10701.4799ecst.

[4] C. Yeh, G. D. Jo, Y.-J. Ko, and H. K. Chung, "Perspectives on 6g wireless communications," *ICT Express*, vol. 9, no. 1, pp. 82–91, 2023, ISSN: 2405-9595. DOI: https://doi.org/10.1016/j.icte.2021.12.017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S240595952100182X.

[5] Y. Zuo, J. Guo, N. Gao, Y. Zhu, S. Jin, and X. Li, "A survey of blockchain and artificial intelligence for 6g wireless communications," *IEEE Communications Surveys Tutorials*, vol. 25, no. 4, pp. 2494–2528, 2023. DOI: 10.1109/COMST.2023.3315374.

[6] N. Kaur and L. Gupta, "Securing the 6g–iot environment: A framework for enhancing transparency in artificial intelligence decision-making through explainable artificial intelligence," *Sensors*, vol. 25, no. 3, 2025, ISSN: 1424-8220. DOI: 10.3390/s25030854. [Online]. Available: https://www.mdpi.com/1424-8220/25/3/854.

[7] Q. Xue et al., "A survey of beam management for mmwave and thz communications towards 6g," *IEEE Communications Surveys Tutorials*, vol. 26, no. 3, pp. 1520–1559, 2024. DOI: 10.1109/COMST.2024.3361991.

[8] S. Liu, X. Yu, R. Guo, Y. Tang, and Z. Zhao, "Thz channel modeling: Consolidating the road to thz communications," *China Communications*, vol. 18, no. 5, pp. 33–49, 2021. DOI: 10.23919/JCC.2021.05.003.

[9] K. Strecker, S. Ekin, and J. F. O'Hara, "Fundamental performance limits on terahertz wireless links imposed by group velocity dispersion," *IEEE Transactions on Terahertz Science and Technology*, vol. 12, no. 1, pp. 87–97, 2022. DOI: 10.1109/TTHZ.2021.3127151.

[10] R. Chataut, M. Nankya, and R. Akl, "6g networks and the ai revolution—exploring technologies, applications, and emerging challenges," *Sensors*, vol. 24, no. 6, 2024, ISSN: 1424-8220. DOI: 10.3390/s24061888. [Online]. Available: https://www.mdpi.com/1424-8220/24/6/1888.

[11] J. M. J. Huttunen, D. Korpi, and M. Honkala, "Deeptx: Deep learning beamforming with channel prediction," *IEEE Transactions on Wireless Communications*, vol. 22, no. 3, pp. 1855–1867, 2023. DOI: 10.1109/TWC.2022.3207055.

[12] S. Wang, M. A. Qureshi, L. Miralles-Pechuán, T. Huynh-The, T. R. Gadekallu, and M. Liyanage, "Explainable ai for 6g use cases: Technical aspects and research challenges," *IEEE Open Journal of the Communications Society*, vol. 5, pp. 2490–2540, 2024. DOI: 10.1109/OJCOMS.2024.3386872.

[13] R. Dwivedi et al., "Explainable ai (xai): Core ideas, techniques, and solutions," *ACM Comput. Surv.*, vol. 55, no. 9, Jan. 2023, ISSN: 0360-0300. DOI: 10.1145/3561048. [Online]. Available: https://doi.org/10.1145/3561048.

[14] İ. Kök, F. Y. Okay, Ö. Muyanlı, and S. Özdemir, "Explainable artificial intelligence (xai) for internet of things: A survey," *IEEE Internet of Things Journal*, vol. 10, no. 16, pp. 14 764–14 779, 2023. DOI: 10.1109/JIOT.2023.3287678.

[15] H. Sun et al., "Advancing 6g: Survey for explainable ai on communications and network slicing," *IEEE Open Journal of the Communications Society*, vol. 6, pp. 1372–1412, 2025. DOI: 10.1109/OJCOMS.2025.3534626.

[16] N. Khan, S. Coleri, A. Abdallah, A. Celik, and A. M. Eltawil, "Explainable and robust artificial intelligence for trustworthy resource management in 6g networks," *IEEE Communications Magazine*, vol. 62, no. 4, pp. 50–56, 2024. DOI: 10.1109/MCOM.001.2300172.

[17] V. Chamola, V. Hassija, A. R. Sulthana, D. Ghosh, D. Dhingra, and B. Sikdar, "A review of trustworthy and explainable artificial intelligence (xai)," *IEEE Access*, vol. 11, pp. 78 994–79 015, 2023. DOI: 10.1109/ACCESS.2023.3294569.

[18] N. Kaur and L. Gupta, *An approach to enhance iot security in 6g networks through explainable ai*, 2024. arXiv: 2410.05310 [cs.CR]. [Online]. Available: https://arxiv.org/abs/2410.05310.

[19] S. Garg, K. Kaur, G. S. Aujla, G. Kaddoum, P. Garigipati, and M. Guizani, "Trusted explainable ai for 6g-enabled edge cloud ecosystem," *IEEE Wireless Communications*, vol. 30, no. 3, pp. 163–170, 2023. DOI: 10.1109/MWC.016.220047.

[20] A. Nechi et al., "Practical trustworthiness model for dnn in dedicated 6g application," in *2023 19th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, 2023, pp. 312–317. DOI: 10.1109/WiMob58348.2023.10187759.

[21] W. Guo, "Explainable artificial intelligence for 6g: Improving trust between human and machine," *IEEE Communications Magazine*, vol. 58, no. 6, pp. 39–45, 2020. DOI: 10.1109/MCOM.001.2000050.

[22] I. Update, "Ericsson mobility report," Ericsson, Stockholm, Sweden, Technical Report, 2013. [Online]. Available: https://images.youmark.it/wp-content/uploads/2013/09/24081357/emr-august-20131.pdf.

[23] X. You, C.-X. Wang, J. Huang, et al., "Towards 6g wireless communication networks: Vision, enabling technologies, and new paradigm shifts," *Science China Information Sciences*, vol. 64, p. 110 301, 2021. DOI: 10.1007/s11432-020-2955-6.

[24] W. Yu et al., *Ai and deep learning for thz ultra-massive mimo: From model-driven approaches to foundation models*, 2025. arXiv: 2412.09839 [eess.SP]. [Online]. Available: https://arxiv.org/abs/2412.09839.

[25] S. Ali et al., "6g white paper on machine learning in wireless communication networks," *CoRR*, vol. abs/2004.13875, 2020. arXiv: 2004.13875. [Online]. Available: https://arxiv.org/abs/2004.13875.

[26] Y. Zhang, M. Alrabeiah, and A. Alkhateeb, "Reinforcement learning of beam codebooks in millimeter wave and terahertz mimo systems," *IEEE Transactions on Communications*, vol. 70, no. 2, pp. 904–919, 2022. DOI: 10.1109/TCOMM.2021.3126856.

[27] M. Alrabeiah and A. Alkhateeb, "Deep learning for mmwave beam and blockage prediction using sub-6 ghz channels," *IEEE Transactions on Communications*, vol. 68, no. 9, pp. 5504–5518, 2020. DOI: 10.1109/TCOMM.2020.3003670.

[28] J. Zhang, G. Zheng, Y. Zhang, I. Krikidis, and K.-K. Wong, "Deep learning based predictive beamforming design," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 6, pp. 8122–8127, 2023. DOI: 10.1109/TVT.2023.3238108.

[29] Z. Hu, Y. Li, and C. Han, "Transfer learning enabled transformer-based generative adversarial networks for modeling and generating terahertz channels," *Communications Engineering*, vol. 3, p. 153, 2024. DOI: 10.1038/s44172-024-00309-x.

[30] A. Nouruzi et al., *Toward a smart resource allocation policy via artificial intelligence in 6g networks: Centralized or decentralized?* 2022. arXiv: 2202.09093 [eess.SP]. [Online]. Available: https://arxiv.org/abs/2202.09093.

[31] A. Patil, S. Iyer, and R. J. Pandya, *A survey of machine learning algorithms for 6g wireless networks*, 2022. arXiv: 2203.08429 [cs.NI]. [Online]. Available: https://arxiv.org/abs/2203.08429.

[32] Q. Wu, S. Zhang, B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface-aided wireless communications: A tutorial," *IEEE Transactions on Communications*, vol. 69, no. 5, pp. 3313–3351, 2021. DOI: 10.1109/TCOMM.2021.3051897.

[33] Y. E. Tok and A. M. Demirtaş, "Optimizing intelligent reflecting surfaces with discrete phase shifts and pilot overhead reduction using deep learning," in *2024 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom)*, 2024, pp. 292–295. DOI: 10.1109/BlackSeaCom61746.2024.10646214.

[34] A. Faisal, I. Al-Nahhal, O. A. Dobre, and T. M. N. Ngatched, "Deep reinforcement learning for ris-assisted fd systems: Single or distributed ris?" *IEEE Communications Letters*, vol. 26, no. 7, pp. 1563–1567, Jul. 2022, ISSN: 2373-7891. DOI: 10.1109/lcomm.2022.3170061. [Online]. Available: http://dx.doi.org/10.1109/LCOMM.2022.3170061.

[35] M. A. S. Sejan, M. H. Rahman, B.-S. Shin, J.-H. Oh, Y.-H. You, and H.-K. Song, "Machine learning for intelligent-reflecting-surface-based wireless communication towards 6g: A review," *Sensors*, vol. 22, no. 14, 2022, ISSN: 1424-8220. DOI: 10.3390/s22145405. [Online]. Available: https://www.mdpi.com/1424-8220/22/14/5405.

[36] R. Kumar and S. Arnon, "Dnn beamforming for leo satellite communication at sub-thz bands," *Electronics*, vol. 11, no. 23, 2022, ISSN: 2079-9292. DOI: 10.3390/electronics11233937. [Online]. Available: https://www.mdpi.com/2079-9292/11/23/3937.

[37] G. Fontanesi et al., "A deep-nn beamforming approach for dual function radar-communication thz uav," *IEEE Transactions on Vehicular Technology*, vol. 74, no. 1, pp. 746–760, 2025. DOI: 10.1109/TVT.2024.3453194.

[38] Y. J. Tan et al., "Self-adaptive deep reinforcement learning for thz beamforming with silicon metasurfaces in 6g communications," *Opt. Express*, vol. 30, no. 15, pp. 27 763–27 779, Jul. 2022. DOI: 10.1364/OE.458823. [Online]. Available: https://opg.optica.org/oe/abstract.cfm?URI=oe-30-15-27763.

[39] W. Yu et al., "An adaptive and robust deep learning framework for thz ultra-massive mimo channel estimation," *IEEE Journal of Selected Topics in Signal Processing*, vol. 17, no. 4, pp. 761–776, 2023. DOI: 10.1109/JSTSP.2023.3282832.

[40] A. M. Elbir, W. Shi, A. K. Papazafeiropoulos, P. Kourtessis, and S. Chatzinotas, "Near-field terahertz communications: Model-based and model-free channel estimation," *IEEE Access*, vol. 11, pp. 36 409–36 420, 2023. DOI: 10.1109/ACCESS.2023.3266297.

[41] Y. Chen and C. Han, "Deep cnn-based spherical-wave channel estimation for terahertz ultra-massive mimo systems," in *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, 2020, pp. 1–6. DOI: 10.1109/GLOBECOM42002.2020.9322174.

[42] A.-A. A. Boulogeorgos et al., *Artificial intelligence empowered multiple access for ultra reliable and low latency thz wireless networks*, 2022. arXiv: 2208.08039 [eess.SP]. [Online]. Available: https://arxiv.org/abs/2208.08039.

[43] Z. Hu, C. Han, Y. Deng, and X. Wang, "Multi-task deep reinforcement learning for terahertz noma resource allocation with hybrid discrete and continuous actions," *IEEE Transactions on Vehicular Technology*, vol. 73, no. 8, pp. 11 647–11 663, 2024. DOI: 10.1109/TVT.2024.3381238.

[44] S. Nie, J. M. Jornet, and I. F. Akyildiz, "Deep-learning-based resource allocation for multi-band communications in cubesat networks," in *2019 IEEE International Conference on Communications Workshops (ICC Workshops)*, 2019, pp. 1–6. DOI: 10.1109/ICCW.2019.8757157.

[45] H. Yang, A. Alphones, Z. Xiong, D. Niyato, J. Zhao, and K. Wu, "Artificial-intelligence-enabled intelligent 6g networks," *IEEE Network*, vol. 34, no. 6, pp. 272–280, 2020. DOI: 10.1109/MNET.011.2000195.

[46] A. Barredo Arrieta et al., "Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai," *Information Fusion*, vol. 58, pp. 82–115, 2020, ISSN: 1566-2535. DOI: https://doi.org/10.1016/j.inffus.2019.12.012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1566253519308103.

[47] F. Rezazadeh, S. Barrachina-Muñoz, E. Zeydan, H. Song, K. Subbalakshmi, and J. Mangues-Bafalluy, "X-grl: An empirical assessment of explainable gnn-drl in b5g/6g networks," in *2023 IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN)*, 2023, pp. 172–174. DOI: 10.1109/NFV-SDN59219.2023.10329778.

[48] S. Lundberg and S.-I. Lee, *A unified approach to interpreting model predictions*, 2017. arXiv: 1705.07874 [cs.AI]. [Online]. Available: https://arxiv.org/abs/1705.07874.

[49] M. T. Ribeiro, S. Singh, and C. Guestrin, ""why should I trust you?": Explaining the predictions of any classifier," *CoRR*, vol. abs/1602.04938, 2016. arXiv: 1602.04938. [Online]. Available: http://arxiv.org/abs/1602.04938.

[50] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 618–626. DOI: 10.1109/ICCV.2017.74.

[51] A. D. Raha, A. Adhikary, G. Mrityunjoy, M. Halder, and C. S. Hong, "Efficient and trustworthy beamforming for 6g: A spatial attention-based deep learning approach," Apr. 2024.

[52] H. Chefer, S. Gur, and L. Wolf, "Transformer interpretability beyond attention visualization," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 782–791. DOI: 10.1109/CVPR46437.2021.00084.

[53] M. Nandan, S. Mitra, and D. De, "Graphxai: A survey of graph neural networks (gnns) for explainable ai (xai)," *Neural Computing and Applications*, 2025. DOI: 10.1007/s00521-025-11054-3.

[54] N. Liu, Q. Feng, and X. Hu, "Interpretability in graph neural networks," in *Graph Neural Networks: Foundations, Frontiers, and Applications*, L. Wu, P. Cui, J. Pei, and L. Zhao, Eds. Singapore: Springer Nature Singapore, 2022, pp. 121–147, ISBN: 978-981-16-6054-2. DOI: 10.1007/978-981-16-6054-2_7. [Online]. Available: https://doi.org/10.1007/978-981-16-6054-2_7.

[55] Z. Ying, D. Bourgeois, J. You, M. Zitnik, and J. Leskovec, "Gnnexplainer: Generating explanations for graph neural networks," *Advances in neural information processing systems*, vol. 32, 2019. [Online]. Available: https://www.researchgate.net/publication/379657688_Efficient_and_Trustworthy_Beamforming_for_6G_A_Spatial_Attention-Based_Deep_Learning_Approach.

[56] H. Yuan, H. Yu, J. Wang, K. Li, and S. Ji, "On explainability of graph neural networks via subgraph explorations," *CoRR*, vol. abs/2102.05152, 2021. arXiv: 2102.05152. [Online]. Available: https://arxiv.org/abs/2102.05152.

[57] D. Luo et al., "Parameterized explainer for graph neural network," *CoRR*, vol. abs/2011.04573, 2020. arXiv: 2011.04573. [Online]. Available: https://arxiv.org/abs/2011.04573.

[58] P. E. Pope, S. Kolouri, M. Rostami, C. E. Martin, and H. Hoffmann, "Explainability methods for graph convolutional neural networks," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 10 764–10 773. DOI: 10.1109/CVPR.2019.01103.

[59] M. Bugueño, R. Biswas, and G. de Melo, "Graph-Based Explainable AI: A Comprehensive Survey," working paper or preprint, Jul. 2024. [Online]. Available: https://hal.science/hal-04660442.

[60] Q. Huang, M. Yamada, Y. Tian, D. Singh, and Y. Chang, "Graphlime: Local interpretable model explanations for graph neural networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 7, pp. 6968–6972, 2023. DOI: 10.1109/TKDE.2022.3187455.

[61] Y. Zhang, D. DeFazio, and A. Ramesh, "Relex: A model-agnostic relational model explainer," *CoRR*, vol. abs/2006.00305, 2020. arXiv: 2006.00305. [Online]. Available: https://arxiv.org/abs/2006.00305.

[62] A. Perotti, P. Bajardi, F. Bonchi, and A. Panisson, *Graphshap: Explaining identity-aware graph classifiers through the language of motifs*, 2023. arXiv: 2202.08815 [cs.LG]. [Online]. Available: https://arxiv.org/abs/2202.08815.

[63] Y. Shi et al., "Machine learning for large-scale optimization in 6g wireless networks," *IEEE Communications Surveys Tutorials*, vol. 25, no. 4, pp. 2088–2132, 2023. DOI: 10.1109/COMST.2023.3300664.

[64] L. Jiao et al., "Advanced deep learning models for 6g: Overview, opportunities, and challenges," *IEEE Access*, vol. 12, pp. 133 245–133 314, 2024. DOI: 10.1109/ACCESS.2024.3418900.

[65] P. Yu et al., "Digital twin driven service self-healing with graph neural networks in 6g edge networks," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 11, pp. 3607–3623, 2023. DOI: 10.1109/JSAC.2023.3310063.

[66] X. Li, M. Chen, Y. Hu, Z. Zhang, D. Liu, and S. Mao, *Jointly optimizing terahertz based sensing and communications in vehicular networks: A dynamic graph neural network approach*, 2024. arXiv: 2403.11102 [cs.NI]. [Online]. Available: https://arxiv.org/abs/2403.11102.

[67] X. Li, M. Chen, Y. Liu, Z. Zhang, D. Liu, and S. Mao, "Graph neural networks for joint communication and sensing optimization in vehicular networks," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 12, pp. 3893–3907, 2023. DOI: 10.1109/JSAC.2023.3322761.

[68] F. Rezazadeh et al., "Toward explainable reasoning in 6g: A proof of concept study on radio resource allocation," *IEEE Open Journal of the Communications Society*, vol. 5, pp. 6239–6260, 2024. DOI: 10.1109/OJCOMS.2024.3466225.

[69] K. Kim, Y. K. Tun, M. S. Munir, W. Saad, and C. S. Hong, "Deep reinforcement learning for channel estimation in ris-aided wireless networks," *IEEE Communications Letters*, vol. 27, no. 8, pp. 2053–2057, 2023. DOI: 10.1109/LCOMM.2023.3280821.

[70] W. Kim, Y. Ahn, J. Kim, and B. Shim, "Towards deep learning-aided wireless channel estimation and channel state information feedback for 6g," *Journal of Communications and Networks*, vol. 25, no. 1, pp. 61–75, 2023. DOI: 10.23919/JCN.2022.000037.

[71] J. Chen et al., "Deep reinforcement learning based resource allocation in multi-uav-aided mec networks," *IEEE Transactions on Communications*, vol. 71, no. 1, pp. 296–309, 2023. DOI: 10.1109/TCOMM.2022.3226193.

[72] D. Yan, B. K. Ng, W. Ke, and C.-T. Lam, "Deep reinforcement learning based resource allocation for network slicing with massive mimo," *IEEE Access*, vol. 11, pp. 75 899–75 911, 2023. DOI: 10.1109/ACCESS.2023.3296851.

[73] Y.-H. Hsu, J.-I. Lee, and F.-M. Xu, "A deep reinforcement learning based routing scheme for leo satellite networks in 6g," in *2023 IEEE Wireless Communications and Networking Conference (WCNC)*, 2023, pp. 1–6. DOI: 10.1109/WCNC55385.2023.10118680.

[74] R. Sutton and A. Barto, "Reinforcement learning: An introduction," *IEEE Transactions on Neural Networks*, vol. 9, no. 5, pp. 1054–1054, 1998. DOI: 10.1109/TNN.1998.712192.

[75] E. Puiutta and E. M. Veith, *Explainable reinforcement learning: A survey*, 2020. arXiv: 2005.06247 [cs.LG]. [Online]. Available: https://arxiv.org/abs/2005.06247.

[76] A. Verma, V. Murali, R. Singh, P. Kohli, and S. Chaudhuri, "Programmatically interpretable reinforcement learning," *CoRR*, vol. abs/1804.02477, 2018. arXiv: 1804.02477. [Online]. Available: http://arxiv.org/abs/1804.02477.

[77] T. Shu, C. Xiong, and R. Socher, "Hierarchical and interpretable skill acquisition in multi-task reinforcement learning," *CoRR*, vol. abs/1712.07294, 2017. arXiv: 1712.07294. [Online]. Available: http://arxiv.org/abs/1712.07294.

[78] G. Liu, O. Schulte, W. Zhu, and Q. Li, "Toward interpretable deep reinforcement learning with linear model u-trees," Dublin, Ireland: Springer-Verlag, 2018, pp. 414–429, ISBN: 978-3-030-10927-1. DOI: 10.1007/978-3-030-10928-8_25. [Online]. Available: https://doi.org/10.1007/978-3-030-10928-8_25.

[79] S. K. Jagatheesaperumal, Q.-V. Pham, R. Ruby, Z. Yang, C. Xu, and Z. Zhang, "Explainable ai over the internet of things (iot): Overview, state-of-the-art and future directions," *IEEE Open Journal of the Communications Society*, vol. 3, pp. 2106–2136, 2022. DOI: 10.1109/OJCOMS.2022.3215676.

[80] Y. Wu, G. Lin, and J. Ge, "Knowledge-powered explainable artificial intelligence for network automation toward 6g," *IEEE Network*, vol. 36, no. 3, pp. 16–23, 2022. DOI: 10.1109/MNET.005.2100541.

[81] B. Kim, Y. Sagduyu, T. Erpek, and S. Ulukus, "Adversarial attacks on deep learning based mmwave beam prediction in 5g and beyond," in *2021 IEEE Statistical Signal Processing Workshop (SSP)*, 2021, pp. 590–594. DOI: 10.1109/SSP49050.2021.9513738.

[82] M. Kuzlu, F. O. Catak, U. Cali, et al., "Adversarial security mitigations of mmwave beamforming prediction models using defensive distillation and adversarial retraining," *International Journal of Information Security*, vol. 22, pp. 319–332, 2023. DOI: 10.1007/s10207-022-00644-0.

[83] Y. Wan, Y. Qu, W. Ni, Y. Xiang, L. Gao, and E. Hossain, "Data and model poisoning backdoor attacks on wireless federated learning, and the defense mechanisms: A comprehensive survey," *IEEE Communications Surveys Tutorials*, vol. 26, no. 3, pp. 1861–1897, 2024. DOI: 10.1109/COMST.2024.3361451.

[84] B. Kim, Y. E. Sagduyu, K. Davaslioglu, T. Erpek, and S. Ulukus, *Over-the-air adversarial attacks on deep learning based modulation classifier over wireless channels*, 2020. arXiv: 2002.02400 [eess.SP]. [Online]. Available: https://arxiv.org/abs/2002.02400.

[85] J. Mao, T. Yin, A. Yener, and M. Liu, *Providing differential privacy for federated learning over wireless: A cross-layer framework*, 2024. arXiv: 2412.04408 [cs.IT]. [Online]. Available: https://arxiv.org/abs/2412.04408.

[86] S. Chen, D. Yu, Y. Zou, J. Yu, and X. Cheng, "Decentralized wireless federated learning with differential privacy," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 9, pp. 6273–6282, 2022. DOI: 10.1109/TII.2022.3145010.

[87] R. Zhou, H. Guo, F. E. C. Teo, and S. Bakiras, "A survey on post-quantum cryptography for 5g/6g communications," in *2023 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI)*, 2023, pp. 1–6. DOI: 10.1109/SOLI60636.2023.10425346.

[88] B. Crook, M. Schluter, and T. Speith, " Revisiting the Performance-Explainability Trade-Off in Explainable Artificial Intelligence (XAI)," in *2023 IEEE 31st International Requirements Engineering Conference Workshops (REW)*, Los Alamitos, CA, USA: IEEE Computer Society, Sep. 2023, pp. 316–324. DOI:

10.1109/REW57809.2023.00060. [Online]. Available: https://doi.ieeecomputersociety.org/10.1109/REW57809.2023.00060.

[89] S. Kruschel, N. Hambauer, S. Weinzierl, S. Zilker, M. Kraus, and P. Zschech, "Challenging the performance-interpretability trade-off: An evaluation of interpretable machine learning models," *Business & Information Systems Engineering*, pp. 1–25, 2025. DOI: https://doi.org/10.1007/s12599-024-00922-2.

[90] A. Assis, J. Dantas, and E. Andrade, "The performance-interpretability trade-off: A comparative study of machine learning models," *Journal of Reliable Intelligent Environments*, vol. 11, no. 1, p. 1, 2025. DOI: https://doi.org/10.1007/s40860-024-00240-0.

[91] S. Mirzaei, H. Mao, R. R. O. Al-Nima, and W. L. Woo, "Explainable ai evaluation: A top-down approach for selecting optimal explanations for black box models," *Information*, vol. 15, no. 1, 2024, ISSN: 2078-2489. DOI: 10.3390/info15010004. [Online]. Available: https://www.mdpi.com/2078-2489/15/1/4.

[92] S. Roy, H. Chergui, and C. Verikoukis, *Towards bridging the fl performance-explainability trade-off: A trustworthy 6g ran slicing use-case*, 2024. arXiv: 2307.12903 [cs.NI]. [Online]. Available: https://arxiv.org/abs/2307.12903.

[93] D. Vale, A. El-Sharif, and M. Ali, "Explainable artificial intelligence (xai) post-hoc explainability methods: Risks and limitations in non-discrimination law," *AI and Ethics*, vol. 2, no. 4, pp. 815–826, 2022. DOI: https://doi.org/10.1007/s43681-022-00142-y.

[94] P. Linardatos, V. Papastefanopoulos, and S. Kotsiantis, "Explainable ai: A review of machine learning interpretability methods," *Entropy*, vol. 23, no. 1, 2021, ISSN: 1099-4300. DOI: 10.3390/e23010018. [Online]. Available: https://www.mdpi.com/1099-4300/23/1/18.

[95] M. H. Alsharif, M. A. M. Albreem, A. A. A. Solyman, and S. Kim, "Toward 6g communication networks: Terahertz frequency challenges and open research issues," *Computers, Materials & Continua*, vol. 66, no. 3, pp. 2831–2842, 2021, ISSN: 1546-2226. DOI: 10.32604/cmc.2021.013176. [Online]. Available: http://www.techscience.com/cmc/v66n3/41062.

[96] H.-J. Song and N. Lee, "Terahertz communications: Challenges in the next decade," *IEEE Transactions on Terahertz Science and Technology*, vol. 12, no. 2, pp. 105–117, 2022. DOI: 10.1109/TTHZ.2021.3128677.

[97] M. Kim, J.-i. Takada, M. Mao, C. C. Kang, X. Du, and A. Ghosh, *Thz channels for short-range mobile networks: Multipath clusters and human body shadowing*, 2024. arXiv: 2412.13967 [eess.SP]. [Online]. Available: https://arxiv.org/abs/2412.13967.

[98] M. Civas and O. B. Akan, *Terahertz wireless communications in space*, 2021. arXiv: 2110.00781 [cs.ET]. [Online]. Available: https://arxiv.org/abs/2110.00781.

[99] A. Ghosh and M. Kim, "Thz channel sounding and modeling techniques: An overview," *IEEE Access*, vol. 11, pp. 17 823–17 856, 2023. DOI: 10.1109/ACCESS.2023.3246161.

[100] M. Wang, Y. Lin, Q. Tian, and G. Si, "Transfer learning promotes 6g wireless communications: Recent advances and future challenges," *IEEE Transactions on Reliability*, vol. 70, no. 2, pp. 790–807, 2021. DOI: 10.1109/TR.2021.3062045.

[101] J. Hall, J. M. Jornet, N. Thawdar, T. Melodia, and F. Restuccia, "Deep learning at the physical layer for adaptive terahertz communications," *IEEE Transactions on Terahertz Science and Technology*, vol. 13, no. 2, pp. 102–112, 2023. DOI: 10.1109/TTHZ.2023.3237697.

[102] A. Shafie, N. Yang, S. A. Alvi, C. Han, S. Durrani, and J. M. Jornet, "Spectrum allocation with adaptive sub-band bandwidth for terahertz communication systems," *IEEE Transactions on Communications*, vol. 70, no. 2, pp. 1407–1422, 2022. DOI: 10.1109/TCOMM.2021.3139887.

[103] K. S. M. H. Ibrahim, Y. F. Huang, A. N. Ahmed, C. H. Koo, and A. El-Shafie, "A review of the hybrid artificial intelligence and optimization modelling of hydrological streamflow forecasting," *Alexandria Engineering Journal*, vol. 61, no. 1, pp. 279–303, 2022, ISSN: 1110-0168. DOI: https://doi.org/10.1016/j.aej.2021.04.100. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S111001682100346X.

[104] F. C. Jong, M. M. Ahmed, W. K. Lau, and H. A. Denis Lee, "A new hybrid artificial intelligence (ai) approach for hydro energy sites selection and integration," *Heliyon*, vol. 8, no. 9, e10638, 2022, ISSN: 2405-8440. DOI: https://doi.org/10.1016/j.heliyon.2022.e10638. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2405844022019260.

[105] I. Molenaar, "Towards hybrid human-ai learning technologies," *European Journal of Education*, vol. 57, no. 4, pp. 632–645, 2022. DOI: 10.1111/ejed.12527.

[106] N. A. Alhaj et al., "Integration of hybrid networks, ai, ultra massive-mimo, thz frequency, and fbmc modulation toward 6g requirements: A review," *IEEE Access*, vol. 12, pp. 483–513, 2024. DOI: 10.1109/ACCESS.2023.3345453.

[107] Y. Hong, J. Wu, and X. Guan, "A survey of joint security-safety for function, information and human in industry 5.0," *Security and Safety*, vol. 4, p. 51, 2025. DOI: https://doi.org/10.1051/sands/2024014.

[108] A. Gupta and A. Nisar, "A novel ai-driven graph-swarm thz slice optimizer for terahertz frequency management and network slicing in 6g/7g oran networks," *International Journal of Communication Systems*, vol. 38, no. 7, e70077, 2025, e70077 IJCS-24-4394.R2. DOI: https://doi.org/10.1002/dac.70077. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/dac.70077. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/dac.70077.

[109] S. M. Lundberg, G. G. Erion, and S.-I. Lee, *Consistent individualized feature attribution for tree ensembles*, 2019. arXiv: 1802.03888 [cs.LG]. [Online]. Available: https://arxiv.org/abs/1802.03888.

[110] V. Ramamoorthi, "Exploring ai-driven cloud-edge orchestration for iot applications," *International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT)*, vol. 9, no. 5, pp. 385–393, Sep. 2023, ISSN: 2456-3307. DOI: 10.32628/CSEIT239072.

[111] H. Wang, B. M. P. Chelvan, M. Golec, S. S. Gill, and S. Uhlig, "Healthedgeai: Gai and xai based healthcare system for sustainable edge ai and cloud computing environments," *Concurrency and Computation: Practice and Experience*, vol. 37, no. 9-11, e70057, 2025. DOI: https://doi.org/10.1002/cpe.70057. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cpe.70057. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/cpe.70057.

[112] R. Alharbi, M. N. Vu, and M. T. Thai, *Learning interpretation with explainable knowledge distillation*, 2021. arXiv: 2111.06945 [cs.LG]. [Online]. Available: https://arxiv.org/abs/2111.06945.

[113] S. C. Ebron, M. Zhang, and K. Yang, *Identifying the truth of global model: A generic solution to defend against byzantine and backdoor attacks in federated learning (full version)*, 2025. arXiv: 2311.10248 [cs.LG]. [Online]. Available: https://arxiv.org/abs/2311.10248.

[114] D. C. Nguyen et al., "Enabling ai in future wireless networks: A data life cycle perspective," *IEEE Communications Surveys Tutorials*, vol. 23, no. 1, pp. 553–595, 2021. DOI: 10.1109/COMST.2020.3024783.

[115] T. Senevirathna, V. H. La, S. Marcha, B. Siniarski, M. Liyanage, and S. Wang, "A survey on xai for 5g and beyond security: Technical aspects, challenges and research directions," *IEEE Communications Surveys Tutorials*, vol. 27, no. 2, pp. 941–973, 2025. DOI: 10.1109/COMST.2024.3437248.

[116] M. N. A. Siddiky, M. E. Rahman, M. S. Uzzal, and H. M. D. Kabir, "A comprehensive exploration of 6g wireless communication technologies," *Computers*, vol. 14, no. 1, 2025, ISSN: 2073-431X. DOI: 10.3390/computers14010015. [Online]. Available: https://www.mdpi.com/2073-431X/14/1/15.

[117] O. T. Basaran and F. Dressler, "Xainomaly: Explainable, interpretable and trustworthy ai for xurllc in 6g open-ran," in *2024 3rd International Conference on 6G Networking (6GNet)*, 2024, pp. 93–101. DOI: 10.1109/6GNet63182.2024.10765734.

[118] S. Roy, H. Chergui, and C. Verikoukis, "Explanation-guided fair federated learning for transparent 6g ran slicing," *IEEE Transactions on Cognitive Communications and Networking*, vol. 10, no. 6, pp. 2269–2281, 2024. DOI: 10.1109/TCCN.2024.3400524.