Received 20 March 2024; accepted 2 April 2024. Date of publication 16 April 2024; date of current version 2 May 2024.

Digital Object Identifier 10.1109/OJCOMS.2024.3386872

Explainable AI for 6G Use Cases: Technical Aspects and Research Challenges

SHEN WANG[1] (Senior Member, IEEE), M. ATIF QURESHI[2], LUIS MIRALLES-PECHUÁN[3], THIEN HUYNH-THE[4] (Senior Member, IEEE), THIPPA REDDY GADEKALLU[5,6] (Senior Member, IEEE), AND MADHUSANKA LIYANAGE[1] (Senior Member, IEEE)

[1]School of Computer Science, University College Dublin, Dublin 4, D04 V1W8 Ireland

[2]ADAPT Centre, Explainable Analytics Group, Faculty of Business, Technological University Dublin, Dublin 2, D02 HW71 Ireland

[3]School of Computing, Technological University Dublin, Dublin 7, D07 EWV4 Ireland

[4]Department of Computer and Communications Engineering, Ho Chi Minh City University of Technology and Education, Ho Chi Minh City 70000, Vietnam

[5]Division of Research and Development, Lovely Professional University, Phagwara 144001, India

[6]Center of Research Impact and Outcome, Chitkara University, Rajpura 140401, India

CORRESPONDING AUTHOR: S. WANG (e-mail: shen.wang@ucd.ie)

This work was supported in part by the European Commission in SPATIAL under Grant 101021808; in part by the Academy of Finland in 6Genesis under Grant 318927; and in part by the Science Foundation Ireland through CONNECT Phase 2 Project under Grant 13/RC/2077_P2, through ADAPT Centre Phase 2 Project under Grant 13/RC/2106_P2, and through Industry Fellowship under Grant 21/IRDIF/9839.

ABSTRACT Around 2020, 5G began its commercialization journey, and discussions about the next-generation networks (such as 6G) emerged. Researchers predict that 6G networks will have higher bandwidth, coverage, reliability, energy efficiency, and lower latency, and will be an integrated "human-centric" network system powered by artificial intelligence (AI). This 6G network will lead to many real-time automated decisions, ranging from network resource allocation to collision avoidance for self-driving cars. However, there is a risk of losing control over decision-making due to the high-speed, data-intensive AI decision-making that may go beyond designers' and users' comprehension. To mitigate this risk, explainable AI (XAI) methods can be used to enhance the transparency of the black-box AI decision-making process. This paper surveys the application of XAI towards the upcoming 6G age, including 6G technologies (such as intelligent radio and zero-touch network management) and 6G use cases (such as industry 5.0). Additionally, the paper summarizes the lessons learned from recent attempts and outlines important research challenges in applying XAI for 6G use cases soon.

INDEX TERMS B5G, 6G, AI, XAI, explainability.

I. INTRODUCTION

THE MOBILE network has been drastically revolutionized in the last few decades. The first-generation mobile network (1G) was introduced in the 1980s, allowing calls to be made from a mobile location instead of a fixed one. The second generation (2G) changed the signal transmitted from analog to digital. It enabled services such as Short Messaging Service (SMS) so that both callers and receivers did not have to be "online" at the same time. The third-generation (3G) mobile network increased the data rate to the level of Mbps, which accelerated access to essential Internet services such as Web browsing. The fourth-generation (4G) integrated with all-IP packet switching networks to provide data rates of up to 1 Gbps, enabling mobile users to access data-intensive services such as TikTok video sharing. The ongoing fifth-generation mobile network (5G) technology supports services such as Enhanced Mobile Broadband (eMBB), Ultra-Reliable Low Latency Communications (URLLC), and Massive Machine-Type Communications (mMTC). 5G enables the Internet of Things (IoT) by increasing device density by 100x with much higher data rates (10x) and latency that is 10x less than 4G.

TABLE 1. Summary of important acronyms.

Acronym	Definition
5/6G	Fifth/Sixth Generation
AI	Artificial Intelligence
A/C/D/RNN	Artificial/Convolutional/Deep/Recurrent Neural Network
A/M/V/XR	Augmented/Mixed/Virtual/eXtended Reality
B5G	Beyond 5G
CAV	Connected Autonomous Vehicle
CV	Computer Vision
DARPA	Defense Advanced Research Projects Agency
DRL	Deep Reinforcement Learning
E2E	End to End
ETSI	European Telecommunications Standards Institute
eMBB	Enhanced Mobile Broadband
GDPR	General Data Protection Regulation
IBN	Intent-Based Networking
ICT	Information and Communications Technology
IoE/T	Internet of Everything/Things
KNN	K-Nearest Neighbors
LIME	Local Interpretable Model-Agnostic Explanations
LRP	Layerwise Relevance Propagation
LSTM	Long Short-Term Memory
mMTC	Massive Machine-Type Communications
ML	Machine Learning
MSCA	Marie Skłodowska-Curie Actions
NFV	Network Function Virtualisation
NLP	Natural Language Processing
PIPL	Personal Information Protection Law
QoE/S/T	Quality of Experience/Service/Trust
RL	Reinforcement Learning
SCS	System Causability Scale
SDN	Software Defined Network
SHAP	Shapley Additive exPlanations
SVM	Support Vector Machines
UAV	Unmanned Aerial Vehicles
URLLC	Ultra-Reliable Low Latency Communications
XAI	eXplainable AI
ZSM	Zero-Touch Service and Network Management

5G networks are being commercialized and deployed worldwide. Many organizations have started planning beyond 5G (B5G) to develop the next generation of wireless cellular networks (6G). While B5G and 6G are often used interchangeably in the literature [1], [2], [3], this paper adopts the term 6G for simplicity. Please also note that since there is yet no wide consensus on the definition of 6G, the term "6G" used in our paper refers more generally to the future networks where "human-centric" AI is widely applied. 6G will further extend the connection coverage by achieving space-air-ground-sea integrated networks [4] to facilitate the Internet of Everything (IoE). Additionally, it will support more data-intensive applications such as full-sensory digital reality. The super reliable and low latency that 6G provides can be well-suited for mission-critical scenarios such as autonomous driving and smart health care.

Given that 5G has a highly softwarized network infrastructure thanks to the software-defined network (SDN) and the network function virtualization (NFV). Building on top of this 5G feature, fully automated network management will be feasible with the power of Artificial Intelligence (AI) in the 6G era to increase the efficiency of network maintenance. More 6G features can be found in various sources such as [5], [6], [7], [8], [9]. Additionally, as first pointed out in [6], we agree that the design of 6G will be "human-centric" rather than "machine-centric". Unlike previous network generations that mainly focused on improving network performance technically, 6G will prioritize implementing a fully automated network powered by AI, such as an intent-based network or intelligent radio. This will satisfy humans' needs without violating personal privacy (e.g., intelligent health and wearable). As a result, for a given time interval, there will be an excessively higher number of AI decisions automatically made due to the high-performance 6G network compared with 5G. The number of incorrect AI decisions is also increasing, which leads to a high risk of the overall AI-based 6G systems. Therefore, such a black-box intelligent 6G system requires promising technologies such as eXplainable AI (XAI) to enhance trust between humans and the network. The role of AI and XAI towards 6G will be discussed further in the following paragraphs.

A. ROLE OF AI FOR 6G

AI will be critical in realizing 6G networks and their applications. There are several ways in which AI can be used in 6G. One of the common ways is through prescriptive, predictive, diagnostic, and descriptive analytics. *Prescriptive analytics* can be used for making decisions or predictions related to edge AI such as cache placement, AI model migration, dynamically scaling network slices and adapting its service function chains, and optimal automatic allocation of resources (e.g., spectrum, cloud, and backhaul).

AI-based *predictive analytics* help to predict the future from the acquired data in real-time for events like resource availability, preference, user behavior, user locations, and traffic patterns, then proactively change the network. Proactive actions can fine-tune the resource allocation, deployment of proactive security solutions, pre-migration of edge services, and edge AI models.

Diagnostic analytics is concerned with detecting faults in the network, thus detecting network anomalies, service impairments, network faults, and the root causes of these network faults, which ultimately helps enhance network security and reliability. Due to the high scalability of the 6G network in terms of users, devices, and services, AI-enabled automatic services are essential for 6G.

Descriptive analytics heavily rely on historical data [10] to enhance the service provider's and network operator's situational awareness. The applications include user perspectives,

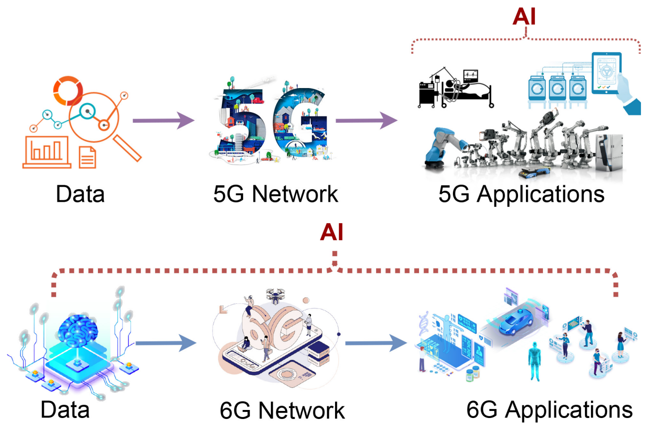


FIGURE 1. An illustration of the role of AI in 5G and 6G Networks. While the focus of AI for 5G will be application-driven, the use of AI in 6G will be aimed at improving the design of all aspects of the network. This includes reliable data sensing, efficient network management, and applications like connected autonomous vehicles.

channel conditions, traffic profiles, network performance, and so on. Furthermore, handling, generating, and processing large volumes of data in real-time and in a collaborative way is yet another complicated task that requires scalable AI. AI will play a vital role in controlling and orchestrating the 6G network. For instance, novel 6G network controlling and orchestrating concepts of Intent-Based Networking (IBN) as well as Zero Touch Service and Network Management (ZSM) are primarily dependent on AI technologies [11]. Novel concepts such as Open RAN or O-RAN will define the development of RAN (Radio Access Network) for future 6G networks. AI is heavily utilized to realize the critical features in O-RAN, such as RAN Intelligent Controller (RIC) frameworks [12].

Figure 1 illustrates the role of AI in 5G and 6G networks. AI will be lightly used in some of the 5G applications. However, AI will be integrated into the E2E (End-to-End) processes in 6G networks.

B. ROLE OF XAI FOR AI-POWERED 6G

XAI is a promising set of technologies that increases the AI black-box models' transparency to explain why certain decisions are made. Especially the high-stakes ones that are made for 6G stakeholders, such as service providers, end-users, and legal auditors. XAI is the key to implementing the "human-centric" AI-powered 6G network.

Fig. 2 shows that AI is integrated into all four layers of the AI-powered 6G network architecture that is proposed in [13]. The first layer, known as the *intelligent sensing layer*, is responsible for gathering data through various sensors, such as phones, watches, drones, or vehicles, in different scenarios, such as space, sea, road, sky, or factory. AI technology can facilitate massive data collection to be a real-time, robust, and scalable process. For instance, this could be done by smartly utilizing the scarce spectrum resources and automatically reporting unreliable data events such as broken sensors. XAI can ensure that the whole process works as expected by providing additional information regarding the AI black-box model. For example, legal auditors may use XAI to check for any privacy violations in the AI training data. Specifically, if the AI decision (e.g., financial credit scoring) is highly biased on a few features that are private personal information such as race, gender, or nationality [14].

The second layer of the AI-powered 6G architecture is the *data mining layer*. Due to the broad coverage of 6G networks, a massive amount of data will be collected from the intelligent sensing layer with a stringent latency requirement. Therefore, the objective of the data mining layer is to perform automatic feature engineering tasks, such as dimension reduction techniques, so that only the most relevant part of the data will be kept for the follow-up processing in the layer of *intelligent control*. This third layer will utilize the filtered data for making decisions such as resource allocations and network management to ensure a certain level of system performance that meets the application requirement. For both the data mining and intelligent control layers, XAI is particularly helpful for service providers to diagnose the root cause of incorrect decisions by AI systems. The top layer of the 6G architecture is the *smart application layer*, which interacts with the end-users who are not technical experts in various scenarios. For example, in the autonomous driving use case, when the AI system suggests turning right, XAI will provide more user-friendly information explaining that the right turn will save five minutes of journey time but can have more curvy roads ahead. Instead of executing decisions straight away, XAI will enhance the trust between stakeholders and the AI-powered 6G networks for prescriptive, predictive, diagnostic, and descriptive analytics.

C. MOTIVATION

Recent research shows the great potential of XAI for computer visions such as medical imaging [15]. However, the challenges when deploying such systems on a large scale (e.g., upcoming 6G systems) remain unclear. The consequences of XAI malfunctioning in some 6G use cases could be significant. Malfunctions may have various causes, including biased spectrum resource allocation [16], inappropriate data collection [17], AI model attack [18], and others. With 6G being deeply coupled with AI in a "full-stack" manner, there is a need for more comprehensive surveys that explore the potential of XAI in implementing AI-enabled human-centric 6G networks.

The Defense Advanced Research Projects Agency (DARPA) launched its XAI program in 2017 [19], which drew the attention of many researchers. Das and Rad [20] compared and analyzed commonly used XAI techniques in terms of their algorithmic mechanisms, taxonomies, and successful applications. Their paper proposed several promising future directions and challenges for XAI. However, existing XAI surveys lack a thorough exploration of the significant potential of XAI in realizing a "human-centric" 6G network. Saad et al. [7] has broadly described the vision

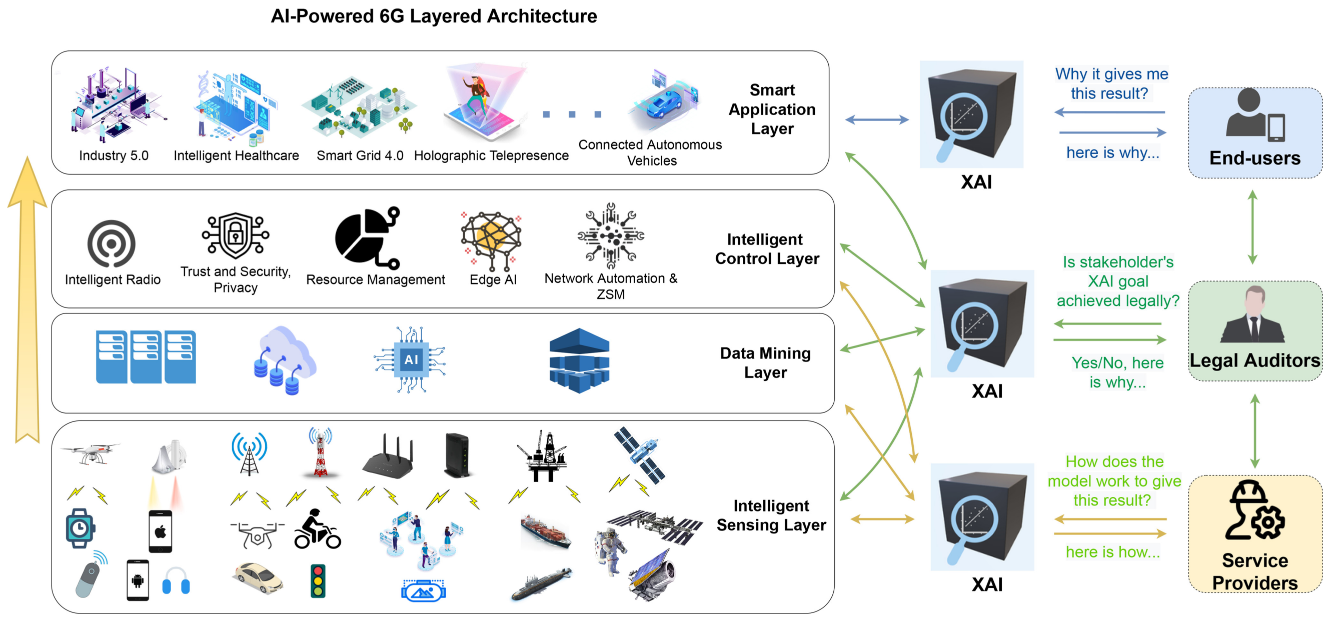


FIGURE 2. An illustration of the benefits (i.e., question-and-answer interactions) of introducing XAI to three typical stakeholders (i.e., end-users, legal auditors, and service providers) across all four layers [13] of AI-powered 6G network. 6G technical aspects discussed in Section IV are illustrated in the intelligent control layer, while some typical 6G use cases discussed in Section V are illustrated at the smart application layer. XAI is built on top of AI so can be deployed on any of these four layers according to specific scenarios and stakeholders.

of 6G, which is far beyond utilizing more spectrum by including more technological trends and driving applications. Morocho-Cayamcela et al. [21] focus on the applications of AI at each main aspect in implementing B5G/6G, ranging from wireless communications to e-health. It also mentioned the trade-off between interpretability and AI algorithms' accuracy but did not extend the discussion on XAI to enhance trust in using 6G cellular systems. Porambage et al. [22] review the recent progress of 6G in security and privacy areas. These areas will likely have many high-stakes decisions by AI systems. However, their contribution lacks discussion of the importance and challenges of XAI for managing the risk of such high-stakes decisions.

In their paper, Guo [16] discussed the potential of XAI in the key enabling technologies, like radio resource management, for 6G at the physical layer and the MAC layer. They also proposed some initial plans for measuring the level of explainability, later formalized as the quality of trust (QoT) in [23] to the users of 6G networks, especially for deep-learning based 6G autonomy. Their paper lacked broader discussions on 6G, especially about the new use cases and the technical aspects that need XAI to uncover the myth of the decision-making process. As mentioned in earlier subsections, there is a high necessity of introducing XAI into AI-powered 6G. Therefore, a comprehensive survey of the state-of-the-art XAI and its potential in building the future 6G networks with a holistic view will be helpful to guide the researchers and practitioners.

Table 2 provides a concise summary comparing important related survey papers. The gap in existing surveys is highlighted, which is the lack of comprehensive analysis of XAI for developing a trustworthy, responsible, and transparent AI-powered 6G network.

D. OUR CONTRIBUTIONS

This paper makes significant contributions, summarized as follows:

- *Bridging the gap between XAI and 6G.* Many existing XAI surveys, such as [20], [24], [25], focus on pure AI applications like natural language processing (NLP) and computer vision (CV). The discussions of XAI to 6G, which is the enabling infrastructure of future AI applications, are unfortunately missing. Similarly, many recent surveys in 6G [7], [21], [22] attempt to cover all possible enabling technologies and applications extensively, without a particular focus on interactions between human and 6G networks, where XAI can play an important role. This survey paper bridges this gap by comprehensively overviewing both XAI and 6G and their connections.
- *A comprehensive survey of XAI to all key aspects of 6G.* In comparison to previous surveys that briefly mentioned XAI's impact on 6G [16], [23], this paper broadens the scope of 6G areas where XAI can contribute. Specifically, this paper goes beyond the smart radio resource management at the physical and MAC layers. Every key 6G technical aspect (e.g., network automation, security, and privacy) and 6G use cases (e.g., industry 5.0, Smart Grid 4.0, Metaverse, and Holographic communication) are examined to investigate how XAI can help in enhancing the transparency and trustworthiness of all 6G stakeholders. The relevant legal frameworks and research projects are

TABLE 2. Summary of important surveys on XAI and 6G.

Ref.	AI	XAI	6G Technical Aspects	6G Use cases	Future Challenges	Remarks
[16]	M	H	H	L	M	A concise survey on the potentials of using XAI for 6G wireless enabling technologies at the physical and MAC layer.
[20]	M	H	L	L	M	A comprehensive survey on the XAI covering commonly used algorithms and applications.
[21]	H	M	H	M	M	A comprehensive survey mainly focuses on the application of machine learning for the future 6G network.
[22]	M	L	H	H	M	A comprehensive survey on the security aspect of 6G using AI solutions
[7]	L	L	H	H	H	A comprehensive survey of 6G visions, technical challenges, use cases, and open research directions.
[23]	M	L	M	M	M	A concise paper on measuring the explainability of deep learning applications in the 6G network automation.
This paper	**H**	**H**	**H**	**H**	**H**	**A comprehensive survey of using XAI for trustworthy and transparent 6G including use cases, requirements/vision, technical aspects, projects, research work, and future research directions.**

H High Coverage (The paper discusses the certain topic in great technical detail with all important sub-topics covered).
M Medium Coverage (The paper discusses the certain topic but not in a great technical detail or missed some important subtopics).
L Low Coverage (The paper never or rarely discusses the certain topic).

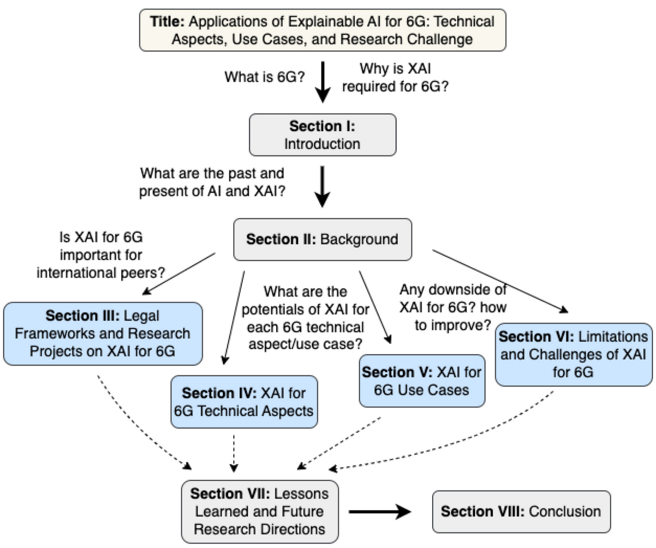


FIGURE 3. Structure and relationships between the sections of the paper.

also reviewed. Moreover, this paper discusses several implementation challenges and possible solutions in applying XAI to 6G.

E. PAPER OUTLINE

The organization of this paper, as shown in Fig. 3, is described as follows. The introduction section outlines the motivation and the overall contribution of this paper, which is followed by the second section that briefly introduces AI and XAI covering their history, technology evolution, popular algorithms, applications, and their trend in 6G areas. Section IV, for each of the six main technical aspects of 6G, namely: intelligent radio, trust and security, privacy, resource management, edge network, and network automation, introduces its motivations, technical requirements, and challenges, and discuss how XAI can improve the level of its trustworthiness. Similarly, Section V discusses each of the six typical 6G use cases, with a particular emphasis on how XAI can help in advancing some of their technical limitations in the 6G age. To demonstrate the importance of our work in the convergence of XAI and 6G, Section III lists several legal frameworks, and ongoing important research projects worldwide about 6G and XAI. Same as many other new technologies, XAI also has its limitations, which are discussed in Section VI, along with its corresponding challenges in the future. Section VII summarises the learned lessons and future research directions for Sections II and IV–VI. Finally, the paper concludes in Section VIII.

II. BACKGROUND

This section provides an essential overview of XAI's background, crucial for understanding its potential in 6G. We discuss the motivation of XAI, concepts relevant to XAI, the taxonomy of XAI algorithms, XAI stakeholders in 6G, and a brief case study on applying XAI to an existing AI-assisted 6G application such as collision avoidance for CAVs.

A. MOTIVATIONS OF XAI

The use of AI algorithms has become increasingly popular. However, one of the main problems with these models, especially the most accurate ones, is that they are considered black-box models. This is because their high internal complexity is difficult to understand. As a result, there is a recent interest in XAI to develop new methods to illustrate how ML models work. XAI will also encourage users to adapt to and trust ML, incorporating it into their work [19], [25].

The relationship between XAI and AI is described as follows: AI can be an independent technology, but XAI cannot exist without AI as it is designed to explain the decisions made by AI. Normally, AI does not always need XAI as not all decisions need to be inspected for debugging systems or legal purposes. However, when explanations are highly required (e.g., high-stake decisions in autonomous driving), XAI can be integrated with all major stages of AI models including data collection (e.g., feature engineering), model training (e.g., a self-explanatory model such as decision tree), and model deployment (e.g., post-hoc XAI model such as LIME). XAI can provide various types of explanations such as text, visuals, rules, linear model weights, feature importance, etc. Developers will collaborate closely with stakeholders to determine the explanation type that makes the most sense for given AI decisions.

B. TYPICAL XAI ALGORITHMS

We briefly introduce some typical XAI algorithms and taxonomies here that are most commonly seen in the literature for 6G or B5G research. We refer survey papers in [20], [24], [25] for more comprehensive studies on XAI methods.

1) MODEL-AGNOSTIC VS MODEL-SPECIFIC

There are two main types of XAI methods: model-agnostic and model-specific. Model-agnostic methods do not consider the internal components of the model, such as its weight and structure parameters, and can therefore be applied to any black-box approach. In contrast, model-specific methods are defined using parameters of the individual model, such as interpreting weights of linear regression or using inferred rules from a decision tree that would be specific to the trained model [26]. There are some advantages of model-agnostic methods [27] such as greater flexibility for developers to choose any ML model for generating interpretation which is different from the actual black-box model that generates decisions.

2) LOCAL VS GLOBAL

Based on the scope of explanations, provided methods can be classified into two classes: local and global methods. Local interpretable methods use a single outcome, or particular prediction or classification results of the model [28] to generate explanations. On the other hand, global methods use the entire inferential ability of the model or overall model behavior [29] to generate explanations. In the local interpretable methods, only specific features and characteristics are essential. For the global methods, feature importance can be used to explain the general behavior of the model.

3) SURROGATE VS VISUAL AID

A way to explain how a black-box model works is by using an interpretable approximate model. This model replaces the black-box model and helps explain how decisions are made. The interpretable approximate model is known as a *surrogate model*. It's trained to make predictions similar to the black-box model. Later, it's used to provide explanations that interpret the decisions made by the black-box model. A black-box model can be a deep neural network (DNN), while decision trees or linear models are examples of interpretable models that can be used as surrogates.

Besides surrogate models, visual explanations aid in generating explanations in a more presentable way showing the inner workings of many model-agnostic. The visual aids can be graphs, scatter plots, heat maps, and so on. For example, ELI5 [30] is a Python library that provides tools to understand machine learning models through visualization and interpretation of their predictions. It supports various ML frameworks and offers functions to explain models and their decisions with weights, highlighting the features contributing to the prediction. In addition, Partial Dependence Plot (PDP) [31] shows the effect of a single or two features on the predicted outcome of a machine learning model, averaged over the joint values of the other features. PDPs help to visualize the relationship between the target response and the features of interest

4) PRE-MODEL, IN-MODEL VS POST-MODEL STRATEGIES

XAI can be applied throughout the entire developmental pipeline of the model. The goal of *pre-modeling* explainability is to describe the dataset to gain better insights into the dataset used to build a model. The main objectives of the pre-model are to perform data summarization, dataset description, perform explainable feature engineering, and conduct exploratory data analysis. Google Facets[1] is an example of pre-model explanations that enable the learning of patterns from large amounts of data.

In contrast, the goal of *in-model* explainability is to develop inherently explainable models instead of generating black-box models. Methodologically, there are different strategies or ways to construct in-model explanations. The most straightforward approach is to adopt an inherently explainable model, such as linear models, decision trees, and rule sets. However, some efforts are needed to generate explanations using these methods, like picking important features. Other approaches are proposed beyond inherently explainable models, such as hybrid models, joint prediction and explanation, and explainability through architectural adjustments.

[1] https://pair-code.github.io/facets/

In the hybrid approach, complex black-box methods are coupled with inherently explainable models to devise a high-performance and explainable model, such as combining a deeply hidden layer of neural network with a KNN model [32]. Also, the model can be trained to provide a prediction and the corresponding explanation jointly [33]. The idea here is to produce a training dataset, where the decision is supplemented with the user's rationale for the decision. Lastly, explanations through architecture adjustments focus on deep network architecture to enhance explainability, such as pushing higher layer filters to represent an object part, as opposed to a mixture of patterns [34]. These approaches within the model have two main shortcomings. Firstly, they assume the availability of explanations in the training dataset, which is often not the case. Secondly, explanations generated by these methods are not necessarily reflective of how model predictions were made, but rather what humans would like to see as an explanation.

The *post-model* explainability method extracts explanations that are inherently not explainable to describe a pre-developed model. These popular post-hoc XAI methods generally operate over four key characteristics: the target, what is to be explained concerning the model; the drivers, what is causing the decision to be explained; the explanation family, how an explanation is going to be presented to a user; and the estimator, the computational process generating the explanation [25].

- *LIME* is a popular model-agnostic XAI strategy [28]. It is a post-hoc algorithm that aims to explain a prediction made by a model. LIME does this by identifying the input features that drive the prediction, assigning importance scores to each feature, and estimating these scores through local perturbations of the input.
  To explain a prediction, LIME creates a surrogate model in the local area. This model is a linear interpretable model that approximates the behavior of the original model in the vicinity of the prediction. By using a local approximation, LIME can work with all types of data, including text, tabular data, images, and graphs. Additionally, it can be used with black-box models, making it a versatile and widely applicable tool in the field of explainable AI.
- *SHAP* calculates feature importance using Shapley values [35], [36] whereas LIME estimates the behavior of a complex model by using a family of interpretable models. Shapley's values are based on cooperative game theory and estimate marginal contribution. Therefore, SHAP generally performs better than LIME. Another advantage of SHAP is that it can explain the global behavior of a model, rather than just a single instance. In general, SHAP is model agnostic. There are also model-specific versions of SHAP to speed up the performance. For example, TreeSHAP is designed for decision trees [37] and Deep SHAP [36] for DNN.
  In contrast to LIME, SHAP takes a more comprehensive approach by calculating feature importance globally for the entire model. This provides a broader perspective on how features affect the model's predictions. Additionally, SHAP values adhere to the consistency property, ensuring that the sum of Shapley values and the baseline align with the model's prediction. Finally, as evaluating all possible permutations can be computationally costly, there are many approaches like Fast Shapley Value Approximations [38] aiming to reduce this computational load.
- *Layerwise Relevance Propagation (LRP)* [39] is an algorithm designed to explain a DNN with an assumption that a classifier can be decomposed into different layers, making it a model-specific method. LRP is designed with the intuition that certain layers of inputs are relevant for the prediction. Activation scores of each neuron are considered through back-pass to identify significant neurons and learn about the input data. LRP is particularly useful for image data, as it highlights meaningful pixels that enable a certain prediction.
  LRP provides detailed insights into neural network decision-making, in contrast to LIME and SHAP, which mainly address global or local feature importance. Nevertheless, LRP's deep network analysis can introduce complexity and computational demands.
- *Grad-CAM* (Gradient-weighted Class Activation Mapping) method was developed in 2017 by Selvaraju et al. [40]. It has gained wide popularity and is used in various fields, especially for images. This method helps create visual explanations for different types of Convolutional Neural Networks (CNNs).
  Grad-CAM algorithm takes the input image and generates a copy of that image but with the relevant pixels highlighted in bright colors like red or yellow. The less relevant areas/pixels are shown with softer colors or are left unchanged. Unlike LRP, which computes the values for the whole network, Grad-CAM only computes the gradients emanating from the CNN's last layer, rendering it adaptable to various CNN architectures [40]. This method helps to figure out what goes wrong in image recognition or why a computer program seems to make random choices. These visual explanations empower users to build trust and confidence in the model's outputs.
- *Counterfactual* is another algorithm that is available for both model-agnostics [41] and model-specific [42] variants. Counterfactual builds on explaining the prediction of the predictor algorithm by finding the slightest change in the input feature values causing the change in the original prediction. For instance, if changing the BMI of the person has flipped the original prediction from illness to being healthy, then using the BMI value is an indicative explanation for correlating with the original prediction. This leads to counterfactual explanations that are easier for humans to understand.

However, there may be multiple possible explanations, making it challenging to determine which is suitable - the simplest or the most complex (i.e., the combination of several features).
- *Bayesian Networks (BN)* are a well-known class of probabilistic models [43]. BNs are based on directed acyclic graphs that compute random variables (nodes) and their relationships (edges) to predict the probability of certain events related to those variables [43]. BNs use Bayesian inference (causality) to estimate such likelihood. Each node of the graph has a probability distribution $P(X_i|\text{Parent}(X_i))$ which represents the conditional probability concerning the parent of that node. The main advantage of BNs is that the graph and the relationships between the variables are interpretable. The predicting reasoning is computed by following the direction of the edges in the graph and the distributions can be visualized. Therefore, BNs are a handy tool for understanding probability distributions, knowledge discovery, and detecting anomalies.
- *Permutation Feature Importance (PI)* [44] measures the increase in a model's prediction error after we permute the feature's values, which breaks the relationship between the feature and the true outcome. This method helps in understanding the influence of each feature on the model's predictions. It differs from LIME as it does not provide local explanations for individual predictions but offers a global view of feature importance by assessing the impact of feature scrambling on model accuracy. Unlike SHAP, it does not rely on time-consuming game-theoretic approaches.

C. TYPES OF EXPLANATIONS BY XAI FOR 6G

Feature importance is one of the most commonly seen types of explanation generated by XAI. It ranks the input data features according to their corresponding contributions to the final output. For example, [45] uses LIME and SHAP to identify the failure in microwave networks from the set of features of link characteristics, G.828 metrics, and power values. Decision tree-based AI models also provide rules as an explanation, as long as the rules are not too complex for stakeholders to understand. For example, a 6-layer decision tree was used for enhancing trust management in network intrusion detection systems.

Saliency maps are well-known explanations for computer vision tasks. They highlight parts of images that lead to the AI output. Saliency maps have great potential in analysing time-series network data such as anomaly detection by highlighting only the key interval that contributes to the prediction results the most. Counterfactual explanations can also be used in facilitating the root cause analysis of future 6G networks by telling the network engineer the minimum possible changes in the values of a certain set of features that could flip the AI algorithm prediction from "unhealthy service" to "healthy service".

TABLE 3. XAI requirements for key stakeholders in 6G.

Stakeholders	Examples	XAI Goals	Preferred Explanation
Service Providers	System designers, data scientists, software developers	To improve 6G AI performance	Logical explanation that can expose the decision-making process
End-Users	Consumers, business partners, policymakers	To trust the 6G AI systems	Easy to understand like visual explanation
Legal Auditors	Solicitors, auditors	To ensure XAI goals are well justified and met	Clear enough explanation to verify if XAI goals can be met

D. XAI STAKEHOLDERS IN 6G

Nearly every sector requires automated algorithmic decision-making, and this demand is evolving into supplementing decisions with explanations generated by the XAI model. With the upcoming 6G making Internet bandwidth faster and available to almost every other device, the demand for AI will be enhanced by XAI within the ecosystem. However, the question remains: who requires XAI, and what level of explanation is deemed reasonable? Also, it is important to note that different stakeholders have different expectations from the explanations [46], and based on the user requirement of XAI [47], stakeholders' demands can be classified broadly into three categories.
- The demand will be useful for *service providers* to help them identify problems or bugs within the system that produce a decision and improve the performance by troubleshooting the decision-making process. Service providers can be system designers, data scientists, AI/XAI researchers, software developers/testers, etc.
- The demand of the *end-users*, who would be interested in understanding the decision for usage and application [48] purpose. For the end-user, the interface of explanations is essential, which should explain the decision in the form of a story that the end-user can easily understand [49]. End-users can be businesses, non-technical people, consumers of technology, and policymakers.
- The demand of the *legal auditors*, who would be interested in auditing legal compliance of automated decision-making algorithm. Here, the legal auditors would look for confirmation that ensures compliance, such as no racial discrimination or gender bias while approving loan applications. These stakeholders can be auditors and other legal professionals.

We summarised the XAI requirements for different 6G stakeholders in Table 3.

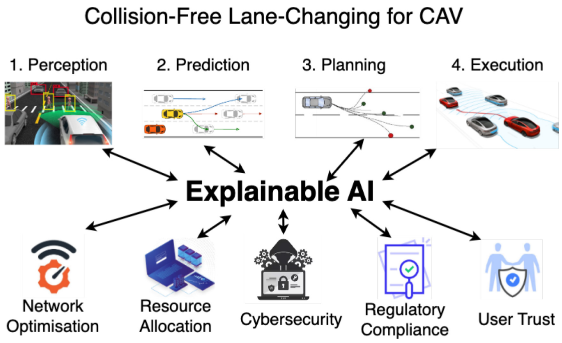


FIGURE 4. An illustration of deploying XAI on 6G using the case study of CAV Collision-Free Lane-Changing.

E. DEPLOYING XAI ON 6G: A CASE STUDY FOR CAV COLLISION-FREE LANE-CHANGING

This subsection presents a case study of deploying XAI solutions on existing AI-enabled 6G applications, focusing on CAV collision-free lane-changing.

1) XAI FOR AI-BASED 6G USE CASE

The emergence of CAVs has brought forth a paradigm shift in automotive technology, requiring sophisticated algorithmic frameworks for safe and efficient operation. Central to this technological revolution is the implementation of collision-free lane-changing mechanisms, which are underpinned by four critical steps: perception, prediction, planning, and execution. At the core of these steps lies the integration of XAI, which is a pivotal advancement ensuring transparency and understandability in AI-driven decision-making processes. XAI plays an indispensable role in each phase, as shown in Fig. 4 beginning with *perception*, where it clarifies the interpretation of sensor data, thus enhancing the reliability of the vehicle's environmental awareness. In the *prediction* phase, XAI demystifies the vehicle's anticipatory capabilities regarding the actions of nearby entities, crucial for accurate maneuver planning. The *planning* stage benefits significantly from XAI by elucidating the rationale behind chosen paths and timings, ensuring safety and regulatory adherence. Lastly, during *execution*, XAI's insights into control system dynamics are instrumental in fine-tuning vehicular responses for optimal maneuver execution. These integrations not only fortify the autonomous system's decision-making acuity but also instill a greater level of trust and accessibility for both developers and end-users.

2) XAI FOR AI-BASED 6G INFRASTRUCTURE

XAI's significance extends beyond the operational mechanics of CAVs, playing a vital role in the development of future 6G networks, which are poised to be the backbone of next-generation vehicular technologies. XAI contributes to the enhancement of 6G networks in several critical areas: network optimization, resource allocation, cybersecurity, regulatory compliance, and user trust, as shown in Fig. 4. In *network optimization*, XAI helps in understanding and improving the AI algorithms responsible for managing network performance, catering specifically to the demanding requirements of CAVs. This is crucial to ensure that the network can support the high bandwidth and low latency needs essential for real-time vehicular communication. Regarding *resource allocation*, XAI ensures transparency in how network resources are distributed, particularly prioritizing the needs of CAVs for safety-critical functions. In the realm of *cybersecurity*, XAI plays a pivotal role in identifying and explaining network anomalies. This is a key factor in safeguarding against cyber threats that could compromise the integrity of autonomous driving systems. *Regulatory compliance* is another area where XAI proves valuable. XAI offers insights necessary for demonstrating that the network aligns with stringent safety and reliability standards required for CAV operations. Lastly, XAI builds *user trust* by making AI operations within the network understandable and accountable, a critical aspect for gaining public acceptance and confidence in these advanced vehicular technologies.

3) SECURING SUCCESSFUL XAI DEPLOYMENT FOR 6G

The successful implementation of XAI in the context of 6G networks supporting CAVs involves a comprehensive and iterative approach. The first step involves a thorough analysis and understanding of AI applications within the autonomous vehicle framework. This analysis should cover the entire spectrum from perception to execution. Doing so enables the identification of specific requirements and challenges that the 6G infrastructure must address to support these applications effectively. Next, the focus should shift towards the development and enhancement of the 6G network infrastructure, ensuring that it is robust enough to handle the demands of CAVs while maintaining transparency and explainability in its operations. The key to success in this endeavor lies in continuously gathering and incorporating feedback from various stakeholders, including network operators, vehicle manufacturers, regulatory bodies, and end-users. This feedback loop is critical for identifying potential areas of improvement in both AI applications and network performance, thereby making the AI-based 6G networks more robust, trustworthy, and aligned with the evolving needs of CAVs.

Furthermore, stakeholder feedback aids in navigating the complex landscape of regulatory compliance, cybersecurity threats, and user trust challenges, ensuring that the network remains not only technologically advanced but also socially acceptable and secure. Lastly, the synergy between XAI and 6G networks, fueled by active stakeholder engagement, paves the way for creating an ecosystem where CAVs operate with unprecedented efficiency, safety, and reliability, marking a significant milestone in the journey towards fully autonomous transportation systems.

III. LEGAL FRAMEWORKS AND RESEARCH PROJECTS ON XAI FOR 6G

This section presents the important legal frameworks and research projects related to the 6G XAI.

A. LEGAL FRAMEWORK FOR EXPLAINABILITY

As explained in Section II, XAI is important for auditors to evolve a legal framework to protect consumer rights under technology usage. Currently, there is no unified law that protects consumer rights for XAI technology. Nevertheless, different regions have started reacting to the evolution of AI and XAI. As we advance towards 6G and XAI, it is anticipated that internationally approved regulations will emerge. For now, we list the adoption of legal frameworks emanating from different regions of the world concerning user privacy and rights to ensure fairness.

- *EU/EEA:* The GDPR [50] is a regulation in EU law on data protection and privacy in the European Union (EU) and the European Economic Area (EEA) and came into effect on 25 May 2018. The GDPR law sets obligations for businesses and grants rights to citizens. Under GDPR, businesses require data protection compliance to ensure data protection concerning users and privacy. Failure to comply can cost up to 20 million euros or 4% of their global revenue. Under GDPR compliance, users have the "right to explanation" in algorithmic decision-making [48], primarily AI systems. In addition, the regulation protects the fair usage of data collection, processing, and application, while maintaining an up-to-date and accurate reflection of data. Finally, it allows users to demand a copy of their data from the business. This regulation comes closest to realizing and facilitating XAI goals of transparency and explanation.
- *USA:* The U.S. has taken a different approach to data protection. Instead of having a general data protection regulation, the U.S. implements sector-specific privacy and data protection policies that work with state laws to protect American citizens' interests. Some of the key sectors are healthcare under HIPAA [51], finance sector and consumer rights under GLBA [52], federal agencies under FISMA [53], and protection of Controlled Unclassified Information in non-federal information systems and organizations under NIST 800-171 [54]. Overall, the U.S. is concerned with data integrity as a commercial asset. In contrast, GDPR gives more to an individual instead of looking at it from the interest of businesses. However, this diversity of legal framework will benefit the adoption of XAI in full spectrum in the 6G world. With this diversity, businesses will communicate with each other through devices and with individuals who would be end-users.
- *Rest of the world:* Ethics, consent, user privacy, law, and transparency are now part of global values. Nearly all countries are bringing forward policies and regulations to ensure their understanding of it to ensure governance, including data-driven decision-making. China developed a new personal data law (PIPL) [55] which came into effect in Nov 2021, drawing its inspiration from GDPR. PIPL tightens how technology giants use data and move private data overseas, with violations resulting in fines up to 5% of the annual revenue of the previous year or CNY50 million. Similarly, PIPL's articles mention automated decision-making related to finance, health, credit status, and more under fair usage and transparency concerning user rights, similar to GDPR. Russia adopted in 2006 a law concerning personal data under Russian Federal Law No. 152-FZ. Even though the law protects individual rights to a certain degree, it is not as comprehensive as the GDPR. Perhaps a revision or expansion of the law can be expected under changing technology and algorithmic decision-making to uphold the spirit of privacy and user protection. Brazil's Lei Geral de Proteçao de Dados (LGPD) [56], Australia's Privacy Amendment (Notifiable Data Breaches) [57], Japan's Act on Protection of Personal Information [58] are steps in similar directions as to those discussed in 2016 for GDPR, which came into effect in 2018.

B. ONGOING REPUTABLE RESEARCH PROJECTS FOR 6G USING XAI

1) EUROPEAN UNION (EU) FUNDED PROJECTS

Due to the popularity of XAI topics, several funding organizations have offered funding for XAI-related topics. European Union (EU) is one of such leading funding organizations that has funded several projects in XAI.

Horizon H2020 (H2020) is one of the biggest funding programs supported by the EU. H2020 is a seven-year funding program that operated from 2014 to 2020 and offered an estimated 80 billion of funding [59]. Under direct H2020 funding, several XAI-related projects were funded, as listed below.

- *AI4EU:* A European AI On Demand Platform and Ecosystem (2019-2021) [60].
- *FeatureCloud* (2019-2024) [61].
- *XMANAI:* Explainable Manufacturing Artificial Intelligence (2020-2024) [62].
- *DEEPCUBE:* Explainable AI Pipelines for Big Copernicus Data (2021-2023) [63].
- *SPATIAL:* Security and Privacy Accountable Technology Innovations, Algorithms, and Machine Learning (2021-2024) [64].
- *STAR:* Safe and Trusted Human Centric Artificial Intelligence in Future Manufacturing Lines (2021-2023) [65].
- *Confidential 6G:* Confidential Computing and Privacy-preserving Technologies for 6G (2023-2025) [66].
- *Rigourous: secuRe desIGn and deplOyment of trUsthwoRthy cOntinUum* computing 6G Services (2023-2025) [66].
- *DAWN4IoE:* Data Aware Wireless Networks for Internet of Everything (2017-2022) [67].
- *Hexa-X-II:* European level 6G Flagship project (2023-2025) [68].

The AI4EU [60] project is building Europe's first AI on-demand platform, which will be used to disseminate AI resources developed by other EU-funded projects. The AI4EU project focuses on XAI and the other interconnected AI domains, such as Collaborative AI, Physical AI, Integrative AI, and Verifiable AI. The FeatureCloud project [61] is focusing on designing secure and trusted medical health systems to reduce the impact of cybercrimes and fuel cross-border collaborative data-mining efforts. To realize this objective, the FeatureCloud project integrates XAI with blockchain and federated learning techniques. The XMANAI project [62] is focusing on the use of XAI for manufacturing to increase trust in AI-based manufacturing processes.

Moreover, the practical utilization of XAI is demonstrated by XMANAI projects in four industrial plants. The DEEPCUBE project [63] is focusing on utilizing XAI Pipelines for extensive data analysis. Primarily, it analyses the Copernicus data, which is collected by the European Union's Copernicus Space Programme [69]. The SPATIAL project [64] is focusing on the development of accountable, resilient, and trustworthy AI-based security and privacy solutions for future networks and ICT systems. Thus, the SPATIAL project focuses on using XAI to ensure the security and privacy of 5G and 6G networks. Several B5G and 6G use cases, such as healthcare and IoT services are considered in this project. The STAR [65] project is studying the use of XAI techniques to increase the transparency of AI-based manufacturing processes and also to improve the user trust level in AI systems.

In addition, H2020 has an element called Marie Skłodowska-Curie actions (MSCAs) [70] which offers grants for all stages of researchers' careers. Under the H2020 MSCA funding, there are two projects for training Early-stage researchers (ESRs) in the domain of XAI applications.

- *NL4XAI:* Interactive Natural Language Technology for Explainable Artificial Intelligence [71]
- *GECKO:* Building greener and more sustainable societies by filling the Knowledge gap in social science and engineering to enable responsible artificial intelligence co-creation [72]

The NL4XAI [71] project focuses on developing self-explanatory XAI systems by utilizing natural language generation and processing, argumentation technology, and interactive technology. The GECKO [72] project is exploring the development of interpretable XAI models to mitigate unintentionally harmful and poorly designed AI models.

2) UNITED STATE GOVERNMENT FUNDED PROJECTS

Defense Advanced Research Projects Agency (DARPA) Information Innovation Office (I2O) in the United States has started a funding program called Explainable Artificial Intelligence (XAI) [73]. Under this program, DARPA has funded several projects focusing on different aspects of XAI:

- *Driving-X:* Study the use of XAI for self-driving vehicles.
- *Rollouts:* Use XAI to establish comfortable human-robot interaction.
- *StarCraft:* Design a self-explaining AI model to play video games.
- *Learning and Communicating Explainable Representations for Analytics and Autonomy:* Design an XAI framework for multi-model analytics and autonomy by recognition, reasoning, and planning domains.
- *COGLE:* Design a system to provide explanations of the learned performance capabilities of an autonomous system.
- *Explainable AI for Assisting Data Scientists:* Study the effectiveness of the XAI system in debugging common ML models.
- *DARE:* Use XAI to improve the accuracy of deep learning models to enable multiple modes of explanation.
- *EQUAS:* Develop a new *Explainable QUestion Answering System* (EQUAS) based on pedagogical and argumentation theories.
- *Model Explanation by Optimal Selection of Teaching Examples:* Analyze the Explanation-by-Examples system to improve the user understanding of black-box ML models.

3) OTHER PROJECTS

In 2017, The Ministry of Science and ICT (MSICT) in South Korea funded the XAI Center [74] which focuses on the research and development of XAI technologies. The XAI center has supported several research activities which are mainly focused on the medical and financial sectors. In 2019, Europe-based Christ-Era organization funded 12 XAI projects under "Explainable Machine Learning-based Artificial Intelligence"[2] funding call. Some of these projects focus on B5G and 6G applications such as digital medicine, robotics, and predictive maintenance.

Although many global-level research activities are being initiated, many of these activities have 6G, 6G technologies, and 6G applications as minor focus. They are still mainly focusing on B5G developments. Table 4 highlights the relevance of these research projects to different aspects of 6G networks.

IV. XAI FOR 6G TECHNICAL ASPECTS

In this section, we discuss the following primary technical aspects of 6G networks: intelligent radio, security, privacy, resource management, edge networks, and network automation. For each technical aspect, we first introduce the background and motivation of its importance in 6G. Then, besides technical requirements, the prospective challenges of the development of regular AI/ML algorithms in wireless networks are analyzed. Finally, we explain how XAI can build trust between humans and AI-enabled machines based on the capability of supporting human-understandable explanations.

[2]https://www.chistera.eu/projects-call-2019

TABLE 4. Important research projects on XAI and their relevance to 6G.

	Funded by	Relevance to 6G	Relevant 6G Technical Aspect						Relevant 6G Applications					
			Intelligent Radio	Trust and Security	Privacy	Resource Management	Edge Network/ EDGE AI	Network Automation, ZSM	Intelligent Health and Wearable	Industry 5.0, Collaborative Robots, Digital Twin	Connected Autonomous Vehicles, UAVs	Smart Grid 2.0	Multi-sensory XR Applications, Holographic Telepresence	Smart Governance
Confidential 6G [66]	EU Horizon Europe	H	✓	✓	✓		✓	✓		✓	✓			
RIGOUROUS [75]	EU Horizon Europe	H		✓	✓	✓	✓	✓	✓	✓				
Hexa-X-II [68]	EU Horizon Europe	H	✓	✓	✓	✓	✓	✓		✓	✓		✓	
SPATIAL [64]	EU H2020	H	✓	✓	✓		✓	✓		✓				
AI4EU [60]	EU H2020	M	✓	✓	✓	✓			✓	✓	✓	✓	✓	✓
FeatureCloud [61]	EU H2020	M		✓	✓	✓			✓					
XMANAI [62]	EU H2020	M	✓	✓	✓	✓	✓			✓	✓	✓	✓	
DEEPCUBE [63]	EU H2020	L		✓	✓	✓			✓		✓			
STAR [65]	EU H2020	M	✓	✓	✓	✓	✓			✓	✓	✓	✓	
NL4XAI [71]	EU MSCA	L		✓	✓				✓				✓	
GECKO [72]	EU MSCA	L		✓	✓	✓						✓		
Driving-X	US DARPA	H		✓	✓						✓			
Rollouts	US DARPA	H		✓	✓			✓		✓				
StarCraft	US DARPA	H		✓	✓	✓							✓	
COGLE	US DARPA	L		✓	✓			✓		✓	✓			
DARE	US DARPA	L				✓								
XAI Center [74]	South Korea MSICT	H	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

L Low Relevance M Medium Relevance H High Relevanc

Notably, for all the subsections "how XAI can help" in this and the subsequent Section V is discussed, considering that not many XAI solutions are built specifically for 6G networks in the existing literature, we have additionally incorporated our analysis on the current AI solutions for the mobile networks and then accordingly share our opinions on how XAI can improve (as well as the potential issues) these solutions (i.e., often for only one layer, as shown in Fig. 2) shortly. We hope this section offers guidance for applying XAI to the future fully AI-powered 6G networks across all layers shown in Fig. 2.

A. INTELLIGENT RADIO
1) INTRODUCTION

The intelligent radio at the intersection of AI and cognitive radio has recently attracted significant attention in solving spectrum problems, including access, monitoring, and management. The rise of modern communication systems with 5G, B5G, and 6G has extended radio services to various industrial domains, exposing several challenging issues and complicated problems in wireless communications. It is feasible to use AI algorithms with automatic learning models to effectively handle channel modeling, intelligent spectrum access, physical layer design, and other network management issues in wireless communications [13], [76]. The emergence of ML, especially DL in the era of big data, has enabled the revealing of essential and unexplored radio characteristics and boosted the progress of wireless and networking technologies with new architectures and novel analysis of pyramid structures.

Over the last decade, we have witnessed the evolution of the traditional black-box base station towards a virtualized next-generation node base (gNB) with the capacity of a functional split, that promotes a new paradigm of Open Radio Access Networks (O-RAN) specialized by disaggregated, virtualized and software-based components, connected over open, programmable, and standardized interfaces with full interoperability across different vendors. In 6G, the AI/ML workflow for intelligent radio is identified by O-RAN with Non-Realtime RAN Intelligent Controller (Non-RT RIC) [77], which should consist of several steps: data collection and processing, training, validation and publishing, development, AI/ML execution and interface, and model maintenance. xApp, as a microservice that is responsible to supervise and manage radio resources through standardized interfaces and service models, is usually designed to control O-RAN slicing policies with real-time responses, in which different AI/ML solutions can be exploited using different key performance measurements, depending on target tasks in the physical layer [78].

2) REQUIREMENTS AND CHALLENGES

Extremely high data rates and low latency of massive machine-type wireless communications are realized as the key requirements of 6G. It can be achieved with an advanced-designed physical layer, wherein several fundamental signal processing and analyzing tasks (e.g., source coding, modulation, orthogonal frequency-division multiplexing (OFDM) modulation, and multi-input multi-output (MIMO) precoding [79]) are powered by AI algorithms [8]. These tasks are typically deployed by the appropriate modules, which follow a forward procedure at the transmitter and an inverse procedure at the receiver. Previously, numerous intelligent radio signal processing approaches were studied with traditional ML algorithms, where expert knowledge of concerning domains is needed to fine-tune radio features and learning models. Being superior to conventional ML, DL has been recently applied to the intelligent radio area to improve performance significantly thanks to its great ability to deal with large, noisy, and confusing raw datasets of radio signals [80]. For example, the accuracy of automatic modulation classification in 5G was improved with CNN architectures while keeping a reasonable complexity [81], [82]. Although DL can extract underlying features from raw radio signals at multi-scale representations to learn complex discrimination patterns, often represents black-box models, lacking interpretability and explainability [83]. Therefore, gaining insight into the rationale behind an AI model's predictions is crucial for aiding network engineers and communication system designers in enhancing system performance and sustainably managing operational risks.

Although it is difficult to find the distinctions in the 6G architecture compared to its predecessor, 5G, it is imperative to focus on certain considerations within the physical layer. Specifically, the integration of AI/ML algorithms should be regarded as essential components facilitating intelligent beamforming, cognitive intelligence-based autonomous radio resource management, intelligent channel coding and modulation, channel estimation, as well as intelligent multiple access and spectrum sharing. Moreover, for future RANs with flexibility, massive interconnectivity, and spectral efficiency, xApps with AI/ML-enabled non-RT RIC should be carried out at the intelligent control layer to optimize radio resource utilization and minimize traffic congestion [84]. A fully user-centric network architecture with ML-driven layers can leverage distributed AI with RAN decisions made by end terminals automatically with no centralized controllers [85], thus enhancing learning efficiency and reducing computing cost if compared with centralized AI.

3) HOW XAI CAN HELP

Despite being superior to traditional ML algorithms (e.g., decision trees, random forest, KNN, ANN, SVM) in terms of accuracy when dealing with large, messy, and confusing practical datasets, DL has a black-box nature that exposes a lack of explainability. For instance, Tunze et al. [86] proposed an advanced automatic modulation classification method with CNN architectures to generally improve accuracy and reduce complexity, however, the method failed to explain why some modulations present better performance than others with the same channel condition and how to predict when the DL model will crash under different practical channel conditions.

In this context, XAI such as LIME, SHAP, and LRP can help 6G stakeholders (e.g., service providers) understand the relations between input data quality and learning efficiency. These tools aid in addressing challenges like imbalance and overfitting, while also enabling the identification of biases within the training and testing datasets. To provide inherent explainability for DL-based modulation, the concept bottleneck model in [87] which comprised a regression network to infer several potential concepts and a classification network to predict the target modulation based on the set of concepts, enables XAI stakeholders to interrogate the classification decisions and address the out-of-training-set problem (i.e., some classes are not seen during training). In [16], Guo pointed out the weak transparency of DL compared with traditional ML for radio resource allocation in the physical layer and MAC layer and then recommended some trustworthy AI techniques to improve explainability. For instance, a case study-based deep feature visualization XAI technique allows the manipulation of key features to optimize a deep model along with different network traffic and channel conditions. Besides, hypothesis testing and

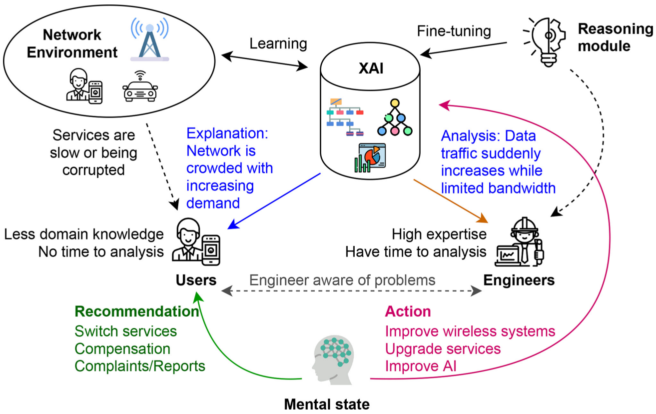


FIGURE 5. An example of XAI for intelligent radio: XAI can explain abnormal phenomena in wireless networks to end-users by ordinary explanation and to system engineers through specialized analysis.

didactic statements with human-machine interfaces are beneficial for elucidating the model learning and decision-making processes, where they play the role of human-like reasoning modules in an XAI integration framework as shown in Fig. 5. For channel modeling with DL, Lee [88] proposed a model-agnostic metamodeling method that can interpret any data-driven channel model into a more understandable form with many transparent mathematical expressions via a symbolic representation technique.Bayesian models are also utilized in [89] to ensure a high explainability of a robust DL based channel estimation for cell-edge users under inter-cell interference.

To achieve intelligent radio in 6G, several new RAN architectures like O-RAN, cloud RAN, virtual RAN, and massive RAN have cooperatively operated and connected in a dense radio environment. XAI will improve the connectivity between mobile users and base stations relying on a set of trustworthy wireless evaluations with dynamic cell selection, intelligent beamforming, channel estimation, adaptive coding and decoding, and automatic modulation recognition. This optimization enhances the efficiency of limited radio spectrum utilization while maintaining a high data transmission rate. Indeed, XAI can help to identify the outliers of radio signal data for training AI models; determine inefficient performing layers and modules in deep neural network architectures for modulation classification and channel estimation; point out which data, model configuration, and training option induce the failure and performance degradation of AI models for intelligent beamforming; and provide additional evaluation metrics to consolidate the AI decision of cell selection in 6G networks. XAI is also realized as the key to the next evolution of xApps to dApps [90] which enables real-time inference and control in O-RAN by cooperatively learning AI models in a distributed manner using locally collected data at RAN nodes. Moreover, XAI can become a sustainable solution to mitigate the difference in hardware configuration and software performance caused by network vendors and operators having different service level agreements, especially, in open network architectures like O-RAN featured by the incorporation of multi-vendor elements, the interoperability and compatibility should be in place. Despite having great potential to revolutionize intelligent radio in 6G with the interpretability and explainability of black-box models, XAI may increase the system's complexity (e.g., where XAI is deployed as an attached module to assist the primary AI model in offering intuitive explanations).

B. TRUST AND SECURITY

1) INTRODUCTION

Since complex 6G networks may contain several heterogeneous dense sub-networks via intelligent connections with cloud-based infrastructures, they will expose some trust and security problems at multiple network connection levels.

6G communication systems should automatically detect proactive threats with intelligent risk mitigation and self-sustaining operations. AI-based trust and security become promising solutions to identify and quickly respond to potential threats automatically [91]. Besides some new threats, 6G networks must cope with existing security issues [92] in previous-generation networks, e.g., SDN, multi-access edge computing (MEC), and NFV. A distributed network that relies on the expansion of device-to-device communication with mesh networks and multi-connectivity is vulnerable and sensitive to the attacks of malicious parties. A hierarchical security protocol can be suitable for wide-area network security and sub-network communication security. Multiple radio units can be attacked over user and control plan microservices at the edge in coexistence scenarios of centralized RAN and distributed core functions. In the perspective of AI-enabled 6G to achieve full automation, ML systems may become the target of several data-based and model-based security threats [93], such as data injection, data manipulation, model evasion, and model modification.

2) REQUIREMENTS AND CHALLENGES

To guarantee high-quality services, the latency criteria of security mechanisms should be taken into consideration in enhanced ultra-reliable, low-latency communications (eURLLC) [22]. Some specialized 6G services with data plane, such as game streaming and remote surgery/telesurgery, require a reliable security solution that can not only effectively defend against cyberattacks but also perform AI/ML-based security analysis promptly with ultra-low latency, to guarantee a certain level of user experience. Moreover, high reliability will demand several extraordinary security solutions to maintain the availability of service operations effectively. High-speed data transmission can reveal certain security issues related to traffic processing security issues, e.g., traffic analysis, AI/ML-related processing flow, and pervasive encryption. These existing issues can be partly handled with distributed security solutions, where raw data and information should be processed locally and on the fly, in decentralized systems, or even in partitioned parts of a network. Distributed ledger technology (with some distinctive features such as transparency, redundancy, and security) and ultra-massive machine type communication (umMTC) can be applied to satisfy security requirements. However, implementing and integrating AI/ML algorithms for resource-constrained devices are still challenging besides the multi-threat analysis of big datasets. Some other issues that may arise from AI-driven security solutions are the responsibility for mistakes made by AI, the scalability and feasibility of AI models in diversified storage and computing infrastructures, and the vulnerability of AI models in distributed systems [94]. For example, the process of uploading local parameter sets from edge devices to a federated center and broadcasting an updated global model to devices can become susceptible to poisoning attacks.

3) HOW XAI CAN HELP

Regarding AI security technologies, transparency in verifying how securely AI systems operate against adversarial machine learning (AML – an attack intentionally fools the AI model by entering deceptive data to make network system unstable, malfunctioning, or unavailable) should be ensured to protect subscribers and mobile communication systems from AML. Besides being created in a reliable system, it is necessary to check whether the AI models operating in user equipment (UE), RAN, and core have been maliciously modified by a malicious attack. O-RAN with an intelligent controller can execute self-healing or recovery procedures if any malicious or abnormal event is detected in AI models. Many recent methods have exploited some advanced ML and DL algorithms to deal with different cybersecurity problems in ORAN and virtualized RAN (vRANs) [93]: poisonous attacks by tampering with the Internet of Everything (IoE) data for training with malicious samples, evasion attacks by injecting disorders and outliers to testing data to circumvent the learned model, API-based attacks to pilfer prediction outcomes, infrastructure physical attacks and communication tampering by interfering communication-computing connections and shutting AI systems down [95]. However, most existing ML/DL-powered security mechanisms cannot explain their final decisions (e.g., how a system achieves threat detection more precisely than attack classification) and response actions (e.g., how an action accordingly responds on time to protect networks from cyberattacks) [96], and consequently cause uncertainty to 6G stakeholders.

XAI presents great potential for improving cybersecurity in 6G IoT networks (i.e., effectively preventing wireless connections, sensory data, intelligent controllers, and applications from different common attacks like poisoning attacks, evasion attacks, physical layer attacks, and model inversion attacks.), ensuring the extreme reliability of IoT-based latency-sensitive services (e.g., industrial automation, emergency response, and remote surgery), and enhancing the trustworthiness of AI-aided security solutions (i.e., AI models have the capability of interpretability and explainability). In [97], Zolanvari et al. introduced transparency relying upon statistical theory (TRUST), a model-agnostic XAI concept that acts as a surrogate explainer to offer multi-level interpretability without sacrificing performance or imposing any restrictions, for cybersecurity in Industrial IoT. By determining the AI's behavior representatives and reasoning on the highest class probability, TRUST provisions transparency of the final decisions made by the AI model. For attack classification, TRUST provides explanations comprehensively for new random samples while presenting high accuracy of over 98% and outperforms LIME in terms of speed and explainability. In [98], an explainable multi-modal hierarchical model (MMHAM) for phishing website detection was proposed to overcome the limited interpretability of deep models. MMHAM leverages a novel shared dictionary learning approach and a hierarchical attention mechanism to align deep representations of fraud cues and facilitate systematic interpretability at different levels, respectively. Although MMHAM can help XAI stakeholders detect phishing websites and proactively react to preventive actions based on explainable insights, it is unable to interpret complicated objects like phishing patterns. In [99], LIME and saliency map XAI methods were applied to interpret AI models developed to detect and classify fingerprinting attacks, in which the most dominant features extracted from raw data are discovered as two post-hoc XAI methods to explain leakage sources in cyber threat intelligence systems. Relying on comprehensive benchmarks with the remove and retrain metric, the two XAI techniques are proficient and compatible with different AI models, including random forest and neural networks. This work also concludes that LIME consumes much more computations than the saliency map in calculating the weights of CNN models. XAI has the potential to offer valuable insights to stakeholders and administrators in trust and security matters. This includes identifying critical features vulnerable to cyberattacks, ranking important features in different AI models against diversified attacks, enhancing detection speed by skipping hidden layers in deep networks, optimizing network configuration and AI training setup for the highest cyberattack classification, and illustrating the relationship between predictive variables and dependent variables in various attack scenarios and system conditions.

C. PRIVACY

1) INTRODUCTION

Data leakage can violate users' privacy, which can be prevented through comprehensive privacy-preserving algorithms. When the number of end devices progressively increases with high data variety, transmitting data over wireless networks, storing data in storage infrastructures, and processing data in computing infrastructures are burdensome with the inserted privacy protection mechanisms [100]. The potential increase in the number of wireless connections in 6G is estimated to be up to 1,000 times greater than that in 5G [101]. Therefore, ensuring high data privacy without service performance degradation is challenging (i.e., obtaining a good trade-off between the enhancement of data privacy and the preservation of service performance in terms of accuracy and processing speed). Moreover, the massive amount of data serving the learning process of statistical AL/ML models will expose a significant challenge for user privacy, which has attracted much more attention from various industrial and academic communities.

2) REQUIREMENTS AND CHALLENGES

With the increasing number of smart devices that enable the effortless collection of massive sensitive data, the data privacy concerns in 6G would be significant due to challenges associated with data transparency [1]. Running intelligent applications on mobile devices and at the edge of the network can become the vulnerable target of privacy attacks. Many problematic privacy concerns will be more severe in the era of big data with 5V features, including volume, velocity, variety, veracity, and value. Adding privacy protection mechanisms will increase communication and computational costs and may not ensure the high quality of 6G-based applications and services unless addressed. Therefore, privacy protection mechanisms should be designed to be cost-efficient besides detecting potential privacy threats automatically and effectively [7]. Addressing privacy concerns is a significant challenge due to the diversity of applications and services. First, easier data acquisition and accessibility can cause regulatory difficulties with data storage, permission, and utilization. Second is the development and integration of AI/ML algorithms into advanced privacy protection mechanisms, which may provoke overloading on resource-constrained devices. Finally, it is noteworthy to balance between the high accuracy of services and the robust protection of user privacy, especially from the perspective of data ownership, access authorization, and other regulations.

3) HOW XAI CAN HELP

Several AI algorithms, including ML and DL, have been applied for privacy protection, wherein privacy threats and attacks are automatically detected and classified to subsequently select the most appropriate privacy-defense mechanism. In [102], a regular ML framework with feature extraction and multi-class classification was developed to detect and assess the privacy risks of mobile phone applications. Despite achieving high classification accuracy of over 90%, some classifiers like SVM and Naïve Bayes presented poor interpretative capabilities, hindering the ability to understand AI decisions and provide meaningful insights for assessment. In [103], a privacy-preserving data mining framework was proposed to identify malicious adversaries that attempt to collect sensitive user data on edge computing platforms. An ensemble learning algorithm using random decision trees was employed to enhance data leakage detection accuracy. However, the insights needed for service providers to plan effective data protection and recovery solutions based on the level of data fragmentation were not considered.

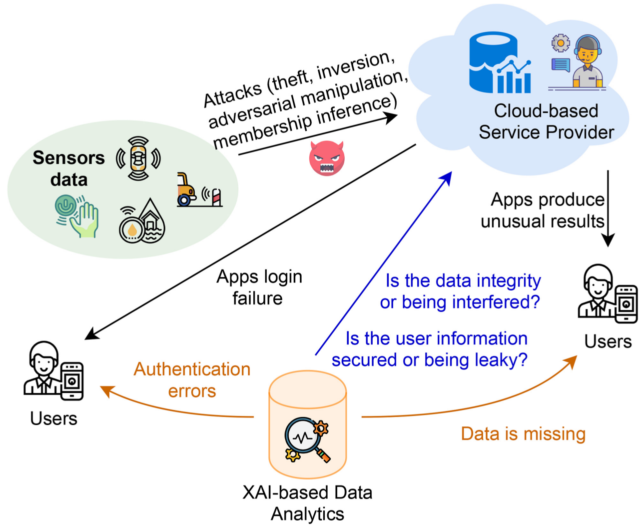


FIGURE 6. An example of XAI for data privacy: XAI-based data analytics can provide specialized analysis and professional explanations to cloud-based service providers regarding undesirable issues of data privacy from end-users.

It is observed that the trustworthiness of AI-aided privacy protection methods is questionable. In this context, some XAI algorithms like LIME and SHAP can contribute as the post-doc models to interpret AI models and expound their decisions by giving a set of explanations representing the impact of each user feature or each data attribute to the final prediction for every single input sample. Furthermore, XAI helps to rank different encryption methods to hide sensitive data, identify which data needs to be shuffled or masked to disassociate its original attributes, and measure the correlations between privacy metrics and data leakage probabilities. Remarkably, some XAI algorithms and methods are beneficial to explainability enhancement of black-box-based privacy models, for example, Deep Learning Important Features (DeepLIFT) finds neurons and weights that significantly affect the final decisions made by DL-aided malicious adversaries detection models, LRP with a set of purposely designed output-oriented propagation rules measures the positive and negative impacts of each layer in DL-based threat detection and classification models, and ProfWeight [104] transfers knowledge from a trained data privacy protection model to other data recovery models and from a data leakage detection model to other secure data storage and integrity models.

Compared with traditional private data release methods that add noise to the original data to improve privacy, several ML and DL algorithms have recently been exploited to generate synthetic fake datasets with little accuracy degradation [105]. Xu et al. [106] proposed GANobfuscator, a differentially generative adversarial network (GAN), to mitigate sensitive information leakage. GANobfuscator can obtain differential privacy by adding specialized noise to the original dataset and adopting gradient pruning techniques during the learning stage. Based on the comprehensive evaluations with different datasets and network architectures, the artificial data generated by GANobfuscator can guarantee the privacy of the original one without information loss. In [107], the synthetic data of electronic health records were generated by a conditional GAN framework to protect sensitive patient information from privacy attacks by adversaries. This GAN framework can prevent patient reidentifiability from statistical measurements of the similarity between the whole dataset and the data of individual patients. Although GAN-based methods are beneficial to privacy protection when generating artificial data to dupe adversaries, they restrain the interpretability and explainability (e.g., reasoning the classification output given by the discriminator to enrich corrective feedback to the generator in a GAN framework). In [108], Nagisetty et al. developed an XAI-guided gradient descent method, denoted XAI-GAN, which can assess the discriminator's decision to explain the feedback passing to the generator in a standard GAN. In particular, XAI-GAN was built as an in-model agnostic explainer to work with different XAI algorithms, such as LIME, DeepSHAP, and saliency maps. Based on the extensive experimental results on different datasets, some key points were concluded: (i) XAI-GAN is better than standard GANs in terms of generated sample quality, (ii) XAI-GAN outperforms standard GANs in terms of classification accuracy when training with the same amount of data, and (iii) XAI-GAN handles the trade-off between data efficiency and training time more effectively than standard GANs.

D. RESOURCE MANAGEMENT

1) INTRODUCTION

Resource management is challenging because of the inherent scarcity of radio resources in wireless communications. It then becomes more difficult in advanced and complicated networks like 6G, wherein the number of smart devices increases rapidly and are involved in different IoT networks (such as cellular IoT, cognitive IoT, and mobile IoT). The overall system performance of a wireless network certainly depends on how it monitors and manages hyper-dimensional resources (e.g., time slots, frequency bands, modulation types, and orthogonal codes) effectively. Besides, the incorporation of wireless channel variations and traffic load attributes in the network design phase impacts connectivity among users having diverse quality-of-service (QoS) requirements. In the context of which new IoT applications demand high data rates, low latency, efficient spectral utilization, and the expansion of personalized IoT services, the problems of resource monitoring and management are crucial [109]. In the last decades, AI/ML algorithms and DL architectures have been widely used to tackle several challenging resource allocation and management issues in 6G-related areas [110]. Concretely, they have contributed to many aspects including massive channel access, power allocation, and interference management, user association and hand-off management, energy management, ultra-reliable and low-latency communication, and heterogeneous QoS. Traditional mechanisms cannot optimize the non-convex problem of resource allocation and are time-consuming to manage resources in a crowded and complicated network. This drawback motivated the discovery of several data-driven ML-based resource allocation and management solutions, in which the high learning capability of AI/ML models is beneficial to the dynamic nature of 6G-enabled IoT networks [111].

2) REQUIREMENTS AND CHALLENGES

Cellular IoT networks in 6G are specialized with extremely high data rates and solid connectivity between heterogeneous devices/users and access points/base stations. The diverse requirements of various IoT services and applications can be met by carefully selecting various network parameters (e.g., channel state information and traffic characteristics) and communication parameters (e.g., angle of arrival and modulation types). These parameters are now remarkably identified by ML and DL in terms of high estimation accuracy and a good ability to deal with big raw data. Some smart devices demand autonomous access to the available spectrum and adaptive tuning of transmission power to mitigate interference and save energy [112]. Some relative system parameters, such as the position and velocity of high-mobility users, propagation conditions, and interference patterns, should be considered when designing an effective ML-based resource allocation solution. Notably, several traditional optimization schemes cannot deal with diversified contexts for integration and cannot respond to varying environments. In numerous IoT applications, ubiquitous and heterogeneous devices have diversified QoS requirements and randomly varying access to network resources, demanding upgrading traditional ML algorithms with RL to fully adapt to the diversity of network requirements and the variation of network conditions [113]. Additionally, the high computational complexity of heuristics-based resource allocation schemes should be considered to be implemented on resource-constrained devices. Compared with 5G, the explosion of new smart terminals in 6G along with a variety of applications in vertical industry markets is pushing mobile network operators (MNOs) to deal with more complex scenarios and deliver more diverse services. So far, the dynamic and diverse demands of 6G users through real-time micro-management of multiple resources, including communication, computing, and storage should be taken into consideration completely. XAI can be deployed and integrated into customized network slicing procedures at MEC servers, wherein the E2E slice consists of several interconnected NFVs from RAN, transport, and core networks, to handle reliably incoming differentiated resource requests from a group of users and reasonably address the demand changes in resource requirements of an individual user.

3) HOW XAI CAN HELP

In the last decades, many ML/DL-based solutions have been introduced for different resource management tasks in 5G

RAN, transport, and core networks, through NFV to decouple software and hardware by virtualizing network functions, including scheduling and duty cycling, resource allocation, power allocation, interference management, resource discovery, cell selection, and mobility estimation. However, the majority of these models lack interpretability and explainability, which hinders the establishment of trust among network operators and end-users.

In 6G, resource management methods should be more reliable and trustworthy with the help of XAI besides ensuring high performance. Specifically, for the network operators, XAI will improve the efficiency of computation offloading, resource allocation, and resource management in a variety of communication networks (including 5G-mmWave communication, O-RAN, long-distance and high-mobility communications (LDHMC), extremely low-power communications (ELPC), and vehicle-to-everything communication) [114] and ensure the primary trained AI models are robust and well-performed with diverse wireless scenarios having various environment agents (e.g., wireless impairments, mobility, and available computing resource). For the end-users, the detailed parameters or explanations of XAI will optimize traffic allocation and minimize the energy consumption of intelligent resource-constrained edge devices, thus enhancing the trust of end-users while offering a high QoS and QoE [115]. Nascita et al. [116] introduced MIMETIC-ENHANCED, a multimodal DL-based mobile traffic classification framework for RAN and transport networks (e.g., V2X - vehicle-to-everything network), in which a general XAI module was deployed to be familiar with different methods, such as additive feature attribution and Deep SHAP. The XAI module in MIMETIC-ENHANCED can typically identify which set of inputs presents the highest confidence probability associated with the model's output. Especially, Deep SHAP allows the classification framework to produce the local explanations based on quantifying the importance value of each input and the global explanations by aggregating the importance values of inputs belonging to each modality. MIMETIC-ENHANCED was experimentally evaluated in terms of trustworthiness (how much XAI stakeholders can trust an estimated confidence entity) and interpretability (the intrinsic reasoning that allows XAI to operate properly). For Internet traffic classification, Callegari et al. [117] built a collection of fuzzy rule-based classifiers using a multi-objective evolutionary learning scheme, in which each classifier plays the role of the XAI classification model and its trade-off between the classification accuracy and the explainability level is optimized individually. Accordingly, based on the input data attributes and the complexity of the problem, XAI stakeholders can select the most appropriate model to achieve a comfortable performance while aptly providing understandable explanations over linguistic IF-THEN rules.

MEC in 6G has seen notable improvements, utilizing ML and DL algorithms to optimize computation offloading. This has resulted in more efficient energy consumption for user equipment. An energy-efficient offloading scheme is proposed in [118] by exploiting the advancements of deep networks to improve the accuracy of multi-component binary classification under various network and user attributes, such as the amount of transmission data, delay, network condition, computational load, etc.

In [119], Baek and Kaddoum deployed a deep recurrent Q-network (DRQN) to cope with a joint request offloading and resource allocation control for heterogeneous services in multifog networks. Compared with deep Q-network and deep convolutional Q-network, DRQN can handle the partial observability problem more effectively thanks to its nodes in recurrent layers to hold internal states and aggregate observations. Many advanced DL-based offloading and resource allocation methods have low transparency and poor explainability which may lead to some risky operation failures and difficulties in fixing or updating deep models. In this context, some XAI candidates (e.g., LIME, LRP, SHAP, and saliency map) can be applied to interpret deep networks (e.g., RL, RNN, and CNN) with the input of time series data [120], which help XAI stakeholders identify which are the most important attributes for different users to optimize offloading computation or minimize energy consumption. By providing insightful analysis and extra relevant information, XAI may tutor stakeholders and system resource managers to comprehend black-box models, such as identifying the least correlative layers in a deep neural network for removal without performance degradation of resource allocation and utilization, determining a set of configurable parameters (e.g., learning rate and regulation factor) in the training stage to achieve a better learning convergence, governing the relationship between accuracy and complexity of resource allocation models to obtain the overall system resource optimization, and pointing out the underlying correlations between global features and local attributes of different architectures to select the most appropriate deep networks.

E. EDGE AI

1) INTRODUCTION

Edge AI is one of the essential components missing in the existing 5G communication networks. Edge AI is a framework that focuses on integrating mobile edge computing, communication networks, and AI seamlessly [121] and is considered to be one of the most critical enabling technologies for the futuristic 6G cellular networks [122]. Recently, many researchers have been working to make the 6G cellular network a fully autonomous and highly intelligent system. Edge AI plays a vital role in realizing human-like intelligence in all the aspects of 6G network systems to improve the quality of experience of the users [123]. AI-enabled decentralized mobile edge servers are deployed at a massive scale for performing processing and decision-making near service requests and data generation. This makes edge AI a vital component, providing accelerated and ubiquitous integration of AI into future 6G networks [124].

2) REQUIREMENTS AND CHALLENGES

Several resources, such as data coordination, model training, caching, and computing, are required to execute AI models in 6G networks [125]. One essential requirement for edge AI in 6G networks is the automatic creation of labels from the massive amount of raw data generated by the wireless devices in the 6G cellular networks. Distributed AI, which performs the computation jointly at the cloud data centers with the distributed edge servers, is one of the key requirements for edge AI. Another essential requirement of edge AI in 6G is personalized AI, through which the decision-making of AI algorithms can be improved by understanding the preferences of the human users [126]. Security is also a crucial requirement for edge infrastructure. Some security threats for edge infrastructure include resource or service manipulation, denial of service attacks, man-in-the-middle attacks, and privacy leakage. AI/ML algorithms can play a significant role in monitoring and predicting security attacks [127], [128]. However, edge AI can vastly improve automation and lower the dependency on human intelligence of 6G cellular networks. In some critical situations, humans have to be involved in decision-making. However, humans may not understand the reason for the predictions of edge AI-based 6G applications, making it very difficult to make confident decisions. One of the key goals of edge AI is self-evolution and self-adaptation so that human efforts can be reduced when processing data and making decisions. Furthermore, the development of the model by dynamically adapting it to unknown events based on the environment and the features of the data is another key goal of edge AI. However, due to the black-box nature of some AI algorithms, it will be challenging to evaluate/audit the effectiveness of AI models for the challenges mentioned earlier. To address these challenges, XAI can be used at the edge of AI in the intelligent control layer so that the 6G will be self-evolutionary and self-adaptive making it fully autonomous and the decisions taken by the AI models can be audited easily by humans.

3) HOW XAI CAN HELP

The lack of explainability can be a severe setback for edge-intelligence-enabled communication networks for some applications, such as vehicle-to-vehicle communications, that require real-time decision-making for preventing crashes, providing driver assistance, and enhancing traffic management. Due to this lack of explainability for edge-intelligence-enabled communication, humans find it difficult to pinpoint the actual origin of problems in catastrophic scenarios. Some factors like human intuition, channel measurements, and theoretical analysis have played a crucial role in the designing of wireless standards and cellular networks. This approach made the domain experts go with either computer simulations or theoretical analysis for validating the building blocks of communication systems. AI models are expected to provide justifications or explainability for cellular networks [129].

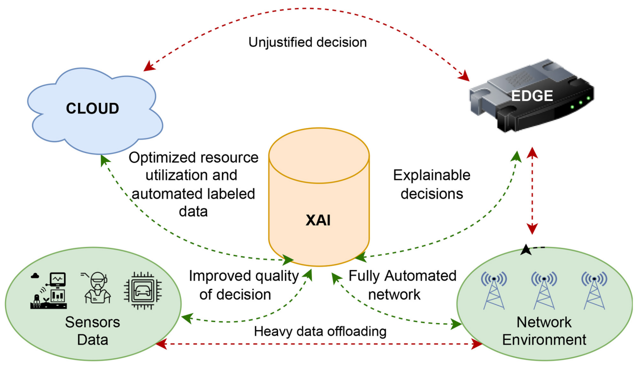


FIGURE 7. An illustration of XAI for edge-assisted 6G: more justified decisions improve the resource allocation between cloud, edge, and sensors.

The base stations integrated with edge intelligence will be granted precise, robust, and high-speed AI algorithms through AI-enabled 6G that will ensure safety-critical mass autonomy for the sub-network settings. The cellular networks are virtually split based on the many services they offer through network slicing. Future AI-enabled 6G network slicing will be allocated based on several human-centric requirements, such as ethics and safety at the sub-network level. XAI can help in explaining the behaviors of mass control systems in terms of the overall policy and also for individual instances, integrated with system performance that can lead to trustworthy supervision of 6G-based services [23].

XAI can play a massive role in guaranteeing the performance of edge AI-enabled 6G networks where several network components are integrated for different requirements through justifiable results. Verification, validation, and auditing of decisions at the edge to address the challenges mentioned above will become simple due to the justification of the decisions from XAI algorithms. Also, humans may have to be involved in some stage of decision-making in mission-critical edges AI-enabled 6G applications like drone-assisted telesurgery systems, smart grid, and border surveillance. In those situations, the job of humans becomes simple as they can easily understand the reasons for the decisions/predictions of the AI algorithms at the edge as depicted in Fig. 7, thus enabling the humans to make better decisions.

The predominant 6G applications, especially involving XR may use AI incorporating Switched Service Networks that can automatically endure high Key Performance Indicators thereby seamlessly managing resources, functions, and network control. The use of AI would enable multi-sensory XR applications to automatically provide energy services to the users to send and develop 3D radio environment maps. The AI-based 6G functions would be complemented by "collective network intelligence" wherein the network intelligence, in association with AI algorithms, would run on the edge devices ensuring distributed autonomy and integration of related services [130]. The adoption of XAI would justify the relevance of the energy services being disseminated to the users in the development of 3D radio environment maps. It would ensure that the most appropriate services are disseminated, being mapped accurately with the specific user requirements. In [131], XAI is proposed to be used by the authors to make the healthcare professionals understand the findings of the DL algorithms to control COVID-19-like pandemics using edge-enabled 5G and beyond networks. The proposed XAI-based solution will ensure that the findings of the ML-based algorithms are justified, which will help the ethical acceptance of the deep neural network model to combat the pandemic situations by healthcare professionals.

In this context, some of the XAI frameworks such as LRP, SHAP, and LIME can be applied to edge AI empowered 6G to interpret the decisions of AI algorithms that can help the stakeholders do not have stringent performance requirements in terms of storage, latency, etc., for autonomous maintenance of the futuristic cellular networks. However, to fully realize the potential of integrating XAI with edge AI empowered 6G, several issues such as a reduced performance of the AI algorithms (to make them more explainable, the complexity of the algorithms may have to be compromised, which may reduce the performance of the AI algorithms), lack of metrics to measure the performance of the XAI algorithms, have to be addressed.

F. NETWORK AUTOMATION AND ZSM

1) INTRODUCTION

New business models will be unlocked using technologies like SDN, MEC, network slicing, and NFV in 5G and beyond cellular networks [132], [133]. The increase in flexibility, cost efficiency, and performance, along with inter-domain cooperation and agility, results in a massive increase in complexity in the management and operation of 5G and beyond networks. Hence, the solutions provided by conventional methods may be inefficient in network and service management. Thus, it is inevitable for management operations through closed-loop automation. Management automation through self-managing capabilities will improve the efficiency and flexibility of delivering services and reduce operating expenses [134]. Zero Touch Network and Service Management (ZSM) was established by ETSI to achieve self-managing capabilities. ZSM reference architecture aims to specify an E2E service and network management services that are fully automated, without the intervention of humans [11], [135].

2) REQUIREMENTS AND CHALLENGES

AI/ML and Big Data analytics are key enablers for 100% automation in cellular networks. AI algorithms can learn from the vast amount of data generated in the 6G cellular network. They can play a vital role in the self-management of the network (self-configuration, self-protecting, self-optimization, and self-healing), resulting in reduced human errors, accelerated time-to-value, and lower operational costs. AI/ML algorithms can be applied to raw data, filter the important events from the large volume of events, identification of problems in the network, and then send the most vital information to the upper layers [136]. However, the successful integration of AI/ML techniques in ZSM for full automation depends on the interpretability and transparency of the AI/ML models to ensure transparency, reliability, trustworthiness, and accountability in AI-enabled ZSM [2]. The type of ML algorithms used and the input data used to train them, which enable the ML algorithms to arrive at the decisions, need to be understood to provide reliable decisions on any automated tasks in ZSM related to delivery, deployment, configuration, assurance, and optimization. The end-user finds it challenging to explain the approach followed by the ML algorithms to arrive at the results due to the increased complexity of the used ML algorithms. This situation is especially faced by users, particularly when a series of updated models are applied to analyze the large volumes of data generated. To address these issues, XAI has to be embedded with the intelligent control layer of 6G architecture, so that the 6G cellular networks are fully automated and the decisions taken by the 6G network based on AI algorithms can be better understood by the network operators.

3) HOW XAI CAN HELP

Emerging networks are becoming more dynamic and complex with each passing day. Hence, data management in contemporary cellular networks is more complex. The traditional approach of data management in cellular networks involves the extraction of information/data generated from measurements/logs and individual events, and then assembling or correlating them together, later creating a summary event that humans can use to make decisions. Due to the huge quantity of data generated, simple correlation techniques for data management don't work in the current scenario. Hence, AI/ML algorithms have a huge role to play in managing present-day cellular networks. However, as more sophisticated ML algorithms are applied, end-users find it challenging to explain the approach used by the ML algorithms, especially in scenarios where multiple updated models are employed over time. With its ability to provide justification and interpretability, XAI has vast potential to address the challenges mentioned above in integrating AI with ZSM. To facilitate XAI for ZSM, the authors in [137] have proposed to add a dictionary that acts as a repository to capture the AI models used in the system. The authors proposed the usage of factor graphs or directed acyclic graphs to represent the taxonomy of AI models. To add the input/output variables and attributes of specific AI/ML algorithms used, the authors also proposed using the algorithm instance repository. In this way, the resulting events can be labeled. The analysts can use this metadata to do reverse engineering and come up with an explanation of the results.

For example, holographic telepresence, a typical 6G application in video conferencing, can project full-motion 3D images in real time. In the case of healthcare, holographic telepresence has immense importance in trauma care and surgeries enabling patients and clinicians to communicate across geographically distributed locations. The need for such technology has been experienced at its peak during the COVID-19 pandemic, which required high-paced care and rendering of specialized services to the patients. The dissemination of such applications is dependent on large bandwidth network communication and related services which would transmit clear holographic transmission using smart devices. The holographic projections are pre-programmed and the necessary projection requires significant manual interventions for efficient resource allocation and communication among the stakeholders [138]. The implementation of ZSM will automate resource allocation and ensure seamless data transfer of holographic communications. ZSM involvement includes collecting full-motion 3D images from the source, which are then analyzed to trigger necessary actions for the executor or consumer of the analytics function. Thus, the intent of the service consumer gets fulfilled autonomously, rendering the required services and resources in 6G. The use of XAI ensures justification for the identification of appropriate managed resources and rendering the most suitable services seamlessly to the stakeholders in a holographic telepresence environment.

Stakeholders can use XAI algorithms such as PIRL as legal auditors and can trace back the automated decisions taken by ZSM to automate several network management services in 5G and beyond networks. Specifically, network administrators can better understand the details regarding network maintenance, implementation of upgrades, and monitoring of attached network devices. In this way, network operators can seamlessly execute performance management, configuration management, and fault management. However, as the decisions taken by ZSM may be critical and may affect the network bandwidth and resource allocation, performance degradation of ML algorithms due to the integration of XAI poses a challenge that has to be addressed.

G. OTHER TECHNICAL ASPECTS

Providing end-to-end virtual networks that cater to the diverse and customized needs of heterogeneous applications is one of the key technical requirements in 6G, that can be realized by network slicing (NS). Managing resources and functions in NS is a challenging and crucial task, where efficient decisions are required at all levels of the network in real-time. Hence, AI can play a vital role in automating the decision-making for these key tasks in NS [139]. As these decisions involve complex management of resources that have financial and service quality, making the human-in-the-loop understand the rationale behind these decisions is of paramount importance. Integrating XAI algorithms into 6G improves the credibility and accountability of resource allocation in NS [140].

V. XAI FOR 6G USE CASES

Most visionary 6G applications need the support of AI and intelligent automation to realize their full capability. This section overviews comprehensively the potentials of XAI for such typical 6G use cases [7], [8], [13], [21], [141] including intelligent health and wearable, industry 5.0, collaborative robots, digital twin, CAVs and UAVs, smart grid 2.0, holographic telepresence, metaverse and smart governance in 6G. Specifically, for each use case, it first introduces the motivation, which explains why this use case is important and required urgently. Then, it analyzes the enabling technologies required, which normally include the 6G communications technologies and AI algorithms. We also introduce some important existing work in the literature under each use case. Last but not least, we explain why XAI can enhance trust between humans and machines in each specific 6G use case and potential issues due to XAI. The XAI requirements of these 6G use cases are summarised in Table 5.

A. INTELLIGENT HEALTH AND WEARABLE, BODY AREA NETWORKS

1) MOTIVATION

The advancement of 6G is expected to drive an innovative development of eHealth systems and improve the performance of medical and healthcare services with advanced AI/ML algorithms [142]. The upcoming 6G communication can provide eHealth applications and services with ultra-high data rates and ultra-low latency for a huge number of connected devices. In this context, an eHealth system is capable of real-time monitoring and tracking, recording health information, and storing eHealth records in cloud-based computing infrastructures. Furthermore, it can exchange medical records and health reports and provide a remote diagnosis by connecting different health service providers in a network [143], [144]. Nowadays, eHealth solutions enabled by 6G can be extended to various scenarios, such as hospitals, sports, homes, and pharmacies, in which the QoS for all eHealth applications and services should be ensured in indoor and outdoor environments. More importantly, 6G-enabled Internet of Medical Things (IoMT) networks promisingly provide precise medical services by applying AI/ML algorithms to process healthcare and medical data besides very high-quality connectivity [145].

2) REQUIREMENTS

As the healthcare data acquired by various multimodal sources has a large volume, high velocity, and diversified variety, exploiting ML algorithms and DL architectures to develop data-driven solutions has been attracting much more attention from healthcare and medical communities via academic research and industrial products. An AI-aided eHealth system should process different data types (e.g., sequential versus high-dimensional data and structured versus unstructured data) and deliver high-quality services with exceptional accuracy (e.g., image-based cancer detection and recognition) and very low latency (e.g., online surgery via video streaming) [146]. In several healthcare and wellness services, raw sensory data collected from wearable devices usually have noise and outliers. Therefore, AI-based solutions should learn complicated patterns from messy datasets effectively. Compared with data in other sectors, healthcare, and medical data are more sensitive to security and privacy, hence ML-based solutions can be studied to automatically detect privacy threats and protect data from cyberattacks in distributed cloud-based systems. Recently, federated learning (FL) was introduced to overcome such kinds of data security and privacy by sharing the information of trained local models instead of the raw data from edge users [147]. Due to the diversity of healthcare data, including sensory data, electronic medical reports, and medical images, collected by different healthcare centers with heterogeneous formats, 6G should specifically regulate data acquisition, storage, and data interoperability in the data mining and analytics layer of 6G-enabled IoMT networks, besides data privacy to prevent the sensitive patient information from cyberattacks. Moreover, from the perspective of AI-aided intelligent healthcare services, a hybrid architecture, i.e., a combination of centralized architecture and decentralized architecture, should be deliberated for distributed learning with a centralized training decentralized execution strategy.

3) EXISTING AI SOLUTIONS

The last few years have witnessed ubiquitous utilization of ML algorithms and DL architectures for a variety of healthcare and medical applications [148], such as physical activity recognition with time-series sensory data and diabetic retinopathy recognition with multimodal images. The authors in [149] proposed an intermediate fusion framework for human activity recognition using sensory data of wearable devices, in which the deep local features extracted by a deep convolutional network were combined with descriptive statistic features to improve the recognition rate. Subsequently, the fused feature vectors were passed into an SVM classifier to predict activities. A comprehensive diabetic retinopathy recognition method [150] leveraged DL technology with CNN architectures to learn the amalgamation between fundus images and wide-field swept-source optical coherence tomography angiography. In this method, a twofold feature augmentation mechanism was advanced to enrich the generalization capacity of the feature level and prevent CNN from the vanishing gradient problem. In another work [151], a two-stage learning model with CNN architecture was presented in a lung nodule detection method to overcome the heterogeneity of lung nodules and the complex pattern of the noisy tomography image dataset. The proposed deep model not only improved the accuracy of early lung cancer detection but also facilitated small-scale datasets with a random mask-based data augmentation scheme. Recently, ML and DL have been applied for many other healthcare and medical applications using different data types, such as sleep analysis with electroencephalography signals collected by wearable in-ear devices [152] and retinopathy risk progression monitoring with electronic medical records [153].

4) HOW XAI CAN HELP

XAI becomes a promising solution for AI models used in the healthcare sector with many benefits such as transparency improvement of AI-based illness diagnosis [154], AI-based drug trial result tracking [155], and model augmentation for health monitoring [156]. Moreover, XAI provides clear explanations on the importance of health parameters such as patient count, patient ages, gender, pre-conditions, and environments to healthcare stakeholders such as drug developers, doctors, health center, etc. [157]. For personalized healthcare in clinical practice, XAI can offer the most appropriate feature set of an intelligent model based on relevant explanations. In [158], an XAI processing pipeline, namely feature selection and classification for improving XAI (SCI-XAI), was developed to automatically select feature engineering as well as classification techniques to achieve the best performance of different fundamental tasks, such as detection and classification. SCI-XAI is benchmarked by a fidelity-to-interpretable ratio that measures how much of the model's interpretability is sacrificed for performance. In [159], Schoonderwoerd et al. proposed a human-centered XAI approach, denoted DoReMi, to generate explanations for clinical decision support systems. Remarkably, this approach can provide explanations with multiple levels of interpretability and understandability for different XAI stakeholders, including service providers and end-users (e.g., clinicians and patients).

In [160], four model-agnostic XAI techniques (including LIME, SHAP, Anchors, and inTrees) were applied to explain the results derived from an XGBoost classifier for ultrasound image analysis in the application of high-risk asymptomatic carotid plaques prediction. The local explanations generated by different XAI techniques are compared and then synthesized to offer global explanations that can explain the operation specification of the entire model. Four XAI techniques were evaluated using different explainability metrics (including clarity, parsimony, completeness, and soundness) in local explanation generation and global explanation synthesis. As a result, model designers conceive the advantages and disadvantages of each model in associating and supporting XGBoost, and further measure the correlations between individual local explanations and the global explanation. With the detailed and clear information derived from XAI, doctors and medical specialists can identify the status of carotid plaques (e.g., shape, location, period, and level) to obtain a more accurate diagnosis and then guide the most appropriate treatment. DL with CNN architectures has been widely used for medical image analysis and demonstrated high performance in terms of accuracy for various fundamental tasks, including detection, classification, and recognition [161]. With the black-box

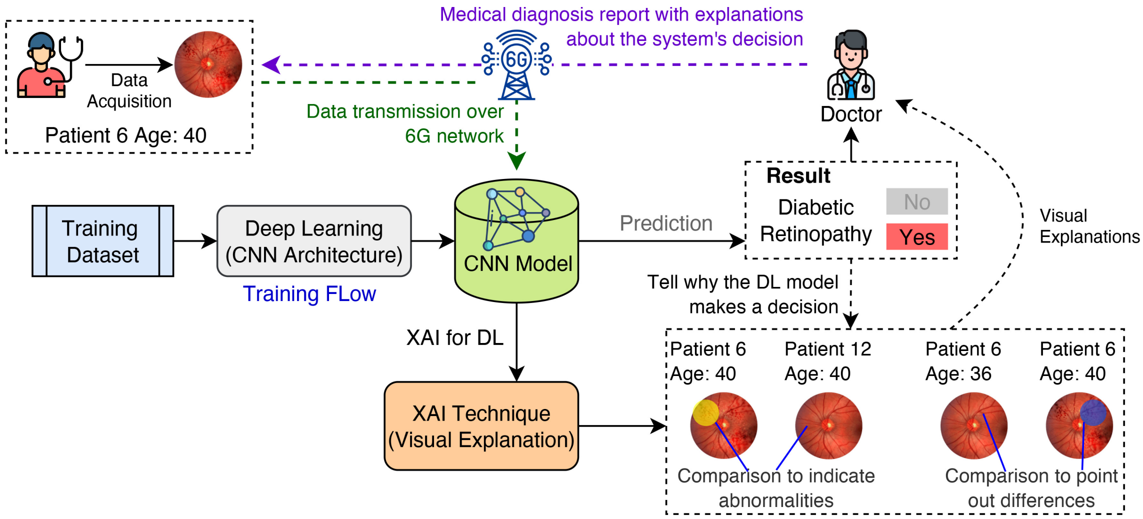


FIGURE 8. An example of XAI with visual explanation for CNN-based diabetic retinopathy recognition.

nature, deep CNN models with a low level of accountability and transparency fail to explain and articulate how their decisions can be reached. Consequently, understanding the operation mechanisms of black-box models is nearly impossible for stakeholders, including end-users and service providers. In this context, it is crucial to develop models that are inherently interpretable to render traceable explanations of AI outcomes. In [162], the combination of saliency maps and gradient class activation mapping (Grad-CAM) [40] was explored to improve DL explanations of prostate lesion localization while keeping high performance in terms of accuracy of lesion classification. The saliency maps and Grad-CAM techniques can emphasize not only the individual pixels but also the feature maps that yield the greatest change in class score, which improves the clarity of visual-based explanations and reduces the amount of noise accordingly in localization and segmentation tasks.

For computer-aided skin lesion analysis and diagnosis with human-friendly explanations, Lucieri et al. [163] proposed ExAID, an explainable AI for a dermatology framework, capable of handling multimodal inputs. The ExAID framework leveraged an activation vector technique to translate the outcomes of deep networks to human-understandable concepts and explored a localization maps technique to highlight these concepts. Subsequently, the relevant concepts are assembled to construct fine textual explanations, which can be combined with concept-wise location information to offer more coherent multimodal explanations. Besides the ability to access data and come to conclusions, XAI can provide doctors and specialists with the decision routine information to understand how those conclusions are reached. Meanwhile, few conclusions require hints of human interpretation [164]. As a result, with XAI, doctors can explain why a certain patient has a high risk of health problems when he should be admitted to the hospital for supervision, and what treatment would be most suitable. An example of applying XAI with a visual explanation technique to recognize image-based diabetic retinopathy is illustrated in Fig. 8. Although current XAI techniques can fulfill the interpretability of black-box models with meaningful explanations to XAI stakeholders, they reveal some serious concerns about the security and privacy of medical data and patient results [165]. In summary, for the service providers, XAI will improve the quality and privacy of healthcare and medical data that is collected in the 6G intelligent sensing layer across heterogeneous IoMT devices and medical systems (e.g., smartphones, smartwatches, electronic medical report systems, and medical imaging systems), enhance the compatibility and interoperability of hardware and software produced by different original equipment manufacturers, and improve the performance of healthcare and medical diagnosis with more coherent and confident treatment plans based on the combination of expert knowledge and explainable result. For the end-users, XAI will provide more detailed numerical analytics and semantic explanations on each medical diagnosis' decision to enhance the trust of patients and the confidence of doctors and medical technicians.

B. INDUSTRY 5.0, COLLABORATIVE ROBOTS, DIGITAL TWIN

1) MOTIVATION

Different from Industry 4.0, which spearheads the explosion of IoT, cognitive computing, big data, and AI over technical interconnectivity and decentralization, Industry 5.0 commits the human touch of business and intelligent systems back into development and production. The primary mission of Industry 5.0 is to create a significant revolution in industrial processes, manufacturing, and business, where problem-solving and creativity-making are the superior objectives instead of replacing repetitive jobs of people with automated robots [166]. In this context, the combination of increasingly powerful machines and well-trained experts motivates effective, safe, and sustainable production, in which highly skilled operators and automated robots can work safely and effectively side-by-side on the same manufacturing role to produce personalized and customized products. Such kinds of robots are known as collaborative robots and should be designed to accomplish different heavy-precision tasks with a high consistency guarantee. Digital twins, recognized as virtual models of the process, product, or service enabling data analysis, system monitoring, and operation and performance assessment via simulations, are promising solutions to optimize business and manufacturer outcomes over managing the entire life cycle of a product [167]. Relying on the comprehensively predictive and descriptive capabilities of digital twins allows customers to comprehend the experience of product functionalities along with operational optimization fully, while manufacturers provide maintenance services to guarantee digital twins are manageable and profitable [168]. In summary, Industry 5.0 is made possible by 6G, offering vast personalized information that surpasses the capabilities of 5G. Similarly, the digital twin will go the extra mile due to 6G, providing ultra-low latency for interactions with real-life objects in a much finer granularity and richer information that is far beyond image, audio, and video. In addition, AI integration in different segments (i.e., RAN, Edge, and Core) of 6G networks is needed to realize better digital twin applications than using AI only at the application level (i.e., 5G does not support network-level AI for the moment). To put it simply, the digital twin will completely characterize a virtual representation of the physical system along with remote sensing, computing, communication, security, and privacy technologies to enable 6G-based IoE services and applications with the capability of automatic planning, analysis, proactive monitoring, control, preventive maintenance, optimization, and business decision making. In the meantime, the digital twin should exploit AI as a native part integrated with not only the smart application layer but also other ones, including the intelligent sensing layer, data mining layer, and intelligent control layer, thus allowing the deployment of different types of digital twins (e.g., from micro to macro level with component/part, asset, system/unit, and process) more accurately and consistently with the physical entity.

2) REQUIREMENTS

In the Industry 5.0 era, we expect to see an intensive upgrade from cyber-physical (i.e., using digital technologies to operate factories to reduce human participation) to human-cyber-physical. Interestingly, as with the foreseeable evolution of collaborative robots and digital twins, AI plays a vital role. This involves processing raw data from sensors, analyzing high-level information, and providing decisions and recommendations automatically. At the center stage of this new revolution, humans should work alongside collaborative robots with the support of digital twin systems, teach them to do tedious, repetitive, and dangerous tasks, and correct them when they make operation mistakes or conduct wrong decisions [169]. Besides the requirement for faster and smarter decision-making in such manufacturing tasks and processes, we desire collaborative robots and digital twins for Industry 5.0 to be more understandable and interpretable, which means, they can explain actions and decisions derived from AI/ML models. The complexity and sophistication of AI-powered automated manufacturing systems rapidly increase to argue that humans cannot understand the ambiguous mechanisms of AI systems, especially when they deliver unpredictable and unexpected decisions [170]. For preparing the incoming wave of Industry 5.0 with mega-factory, collaborative robots, and digital twins, 6G should support offloading of real-time intensive computations tasks, hyper-fast data rate, extremely low latency communications, and highly flexible compatibility of massive IoT devices. In addition, as expected to become an essential requirement in the manufacturing environments, the trustworthiness should be improved with secure-by-design for trusted hardware/software and XAI for preventing adversarial ML.

3) EXISTING AI SOLUTIONS

Nowadays, collaborative robots are being developed to support automatic inspection and corrective action in high-precision control systems [171], [172], in which there is an AI-enabled intelligent module designed with DRL to effectively learn and adaptively act based on inspection results. The approach shows two primary advantages: first is learning the AI model continuously without shutting systems down, and second, its extensibility to different real-world scenarios. In [173], a dual-input DL model was developed to improve the performance of human-robot collaboration and allow robots to learn from human demonstrations effectively. This model synthesized the assembly contexts of multiple human demonstration processes and tasks to accomplish suitable assistant actions. Compared with traditional feature-based approaches being more complex and time-consuming for labeling a huge amount of data, the proposed method can annotate data labels automatically by perceiving human demonstrations. In another work [174], RL with CNN architecture was leveraged to optimize the working sequence of human-robot collaborative assembly, which increases the working performance of smart manufacturing systems. Some complicated learning use cases, such as robot random failure and human behavior uncertainties, were further taken into consideration to satisfy real-world conditions. For the promotion of fully smart manufacturing, cognitive digital twins [168] were presented by incorporating different modern digital technologies, including industrial IoT, big data, ML, and virtual reality, which aimed to analyze and simulate operation modules, assets, systems, and processes. In [175], a secure industrial automation system was built based on a digital twin replication model to identify and verify multi-level design-driving security requirements using sophisticated simulations and optimizations. The digital twin technology was also exploited in some heavy industry sectors [176], such as shipbuilding, steel production, and oil

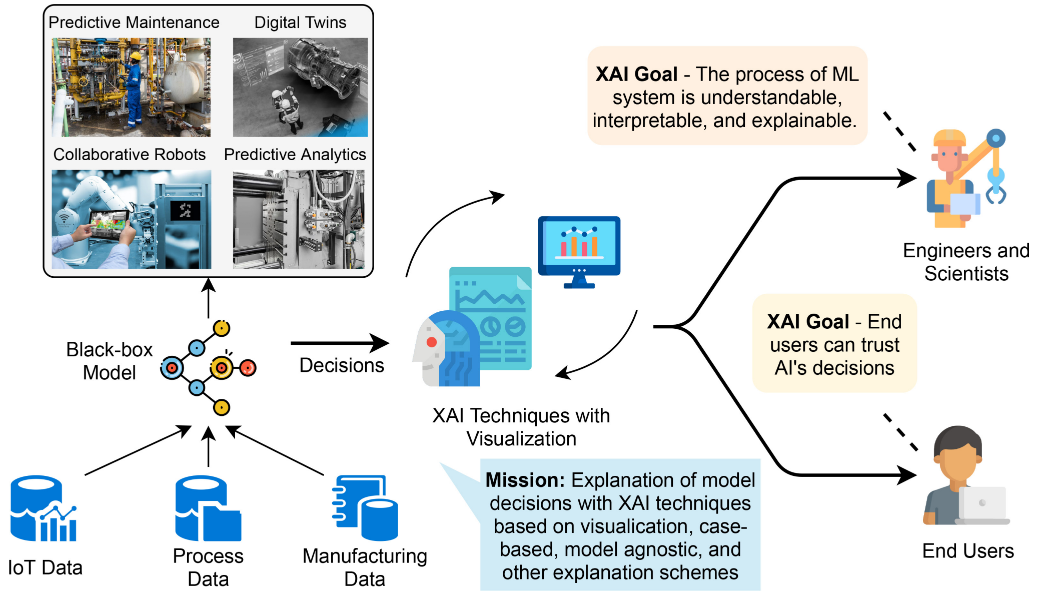


FIGURE 9. An illustration of XAI for Industry 5.0, where the goal of XAI for different subjects varies: XAI helps end users trust AU's decision while XAI makes engineers or scientists understand the process of ML systems completely.

and gas, to enhance safety and productivity while reducing operational costs and minimizing health risks and accidents.

4) HOW XAI CAN HELP

XAI enables humans to understand certain aspects of AI-aided processes or systems, in which XAI answers such kinds of questions: why is the prediction reliable, what are the stable working conditions of an AI model, and when is it likely to crash; and provides more extra deep-analyzed information about system parameters, operation factors, manufacturing environments, and customer feedback to industry stakeholders like system engineers, operators, and factory managers. XAI becomes a sustainable solution for conventional and advanced AI models developed for applications and services in the manufacturing sector for trustworthiness improvement, transparency enhancement, and result interaction of AI-based predictive maintenance [177], ML-based manufacturing diagnosis [178], and data-driven model-based human-robot collaborative assembly [179].

6G with massive URLLC and O-RAN can lead to remarkable advances in Industrial IoT (IIoT) and machine-to-machine (M2M) communications, wherein a huge number of automated connected edge devices and modules in smart factories requires a flexible interface, reliable and low-latency wireless communication, energy efficiency, and split computing with data synchronization, especially comprehensive XAI with the capability of self-detecting abnormality, self-monitoring operation, and predictive maintenance. From the viewpoint of system operators, XAI will improve the efficiency of automatic system diagnosis and maintenance based on the sensory data (e.g., sound, image, vibration signal, acceleration signal, etc.) collected in the 6G intelligent sensing layer, thus reducing time and cost of periodic system maintenance. In [177], a CNN-based bearing faults diagnosis method was proposed to classify vibration signals with the short-time Fourier transform, in which Grad-CAM was applied to generate the model's awareness. By analyzing the awareness, visual explanations of the damage conditions of rolling element bearings can be carried out.

In experiments, the explanations generated by Grad-CAM to articulate the decision of CNN were verified by neural networks, decision trees, and adaptive network-based fuzzy inference systems in terms of correctness and persuasiveness. In [178], a LIME-based XAI framework for chiller fault detection and diagnosis systems, denoted XAI-FDD, was proposed to interpret data-driven classification and regression models like XGBoost and other traditional models. In association with AI models, LIME ranks the importance of sensors based on the failure detection rate, identifies the top of dominant handcrafted features for failure type classification, and estimates the correlation between configurable parameters (e.g., minimum of instance weight and maximum delta step) and learning convergence. Moreover, XAI-FDD provides AI stakeholders with more meaningful diagnosis information, such as a set of frequently fault modules and components in a chiller system and a set of promisingly effective solutions to overcome failures, thus reducing the cost of system maintenance and avoid the system shutting down. In industrial control systems, analysts should intervene and respond immediately when any sensor fault is automatically detected by AI models. However, many high-performance detection models have been limited by weak interpretability and explainability. In [180], a comprehensive XAI framework was deployed with Deep SHAP and Gradient SHAP to explain the prediction results derived from LSTM networks and to interpret the feature importance of predictions without compromising the accuracy of deep models. In experiments, it was concluded that Deep SHAP consumes much processing time to interpret complex deep networks (e.g., RNN and LSTM) and Gradient SHAP presents some advantages in handling multiple input models with a comfortable complexity.

Intelligent robots without explanatory capability may cause failures and dangerous actions unexpectedly, hence interpretability and explainability, including post-hoc rationalization and introspection, should be supposed to ensure no conflict between supplementary explanations and regular requirements of applications. To reach human-like communication in human-robot communications, a novel XAI framework [181] was developed for collaborative robots by constructing a hierarchical AI model with a human-learned And-Or graph-based explanation scheme. A spatial-temporal-causal And-Or graph (STC-AoG) was designed to encode tasks and sub-tasks. It incorporates temporal, spatial, and causal relations between robots and an agent, enabling robots to understand human intentions. Based on the comparison between robot and human mental states updated and inferred by STC-AoG, it is possible to estimate the difference between expected human action and predicted robot response to decide whether explanations are given. Besides graph-based techniques, reinforcement meta-learning XAI algorithms were studied in [182] to interpret spatial-temporal attention and emotion of humanoid robots in collaborative tasks with agents, wherein the explanations are generated and synthesized by a deep belief network, rule-based fuzzy cognitive map, and genetic algorithm to justify robot's self-awareness and decision-making. In the next industrial revolution, digital twins using explainable models and explanation interfaces can become a promising solution to enhance the trustworthiness of AI [179], where 6G will uplift digital twins in many aspects (e.g., effectively handle the massive volume of heterogeneous IoT data in real-time with ultra-low latency and ultra-high throughput, increase flexibility with IIoT deployment and configuration for digital twin refinement, improve analysis accuracy in manufacturing operation and reduce time and cost with predictive maintenance). Interestingly, as an exact virtual representation, or virtual clone/mirror, of a real-world asset, a digital twin enables XAI stakeholders to heighten the explanatory capability with augmented reality and virtual reality technologies, for example, the remaining life cycles of modules/components in a production line and the data transmission failure in a manufacturing network. In summary, XAI enhances the efficiency of AI-based manufacturing operations, particularly with collaborative robots on a production line. Here, 6G facilitates nearly real-time multi-robot task assignment (i.e., factory throughput estimation, device-edge split computing, adaptive computation offloading, optimal energy allocation) in complicated multi-agent environments and supports conventional AI, including ML and DL models, with more relevant information (e.g., useless features/characteristics of AI for removal without performance degradation and useful kernels/layers/modules of DL to improve the accuracy of robot's decision-making). With the end-users, XAI plays the role of providing more detailed explanations and vindications on each action decision and more numerical analytics (simulation, experiment, and test of digital twins before implementing on the physical operation), thus enhancing the trust of customers, operators, and product line engineers.

C. CONNECTED AUTONOMOUS VEHICLES, UAVS

1) MOTIVATION

Recently, the wealthy development of connected and automated vehicle technologies is expected to positively change the manner vehicles move and the approach travelers obtain mobility. Connected autonomous vehicle (CAV) is recognized as one of the important vertical industries in 6G, offering various on-demand services with different quality levels [183]. CAV can mean autonomous vehicles that have the capabilities of connecting other vehicles/infrastructures over wireless communications and sensing the driving environment to achieve safe transportation with little or no human involvement. By incorporating advanced technologies, autonomous vehicles collaborate directly over an intermediate infrastructure to improve the performance and efficiency of smart transportation systems if compared with individual autonomous vehicles without collaborative mechanisms. Indeed, the connected vehicle and automated vehicle technologies should be developed in parallel and closely cooperated to put forward completely smart transportation in the future [184]. To this end, besides 6G-enabled wireless communications, AI plays one of the most important core technologies in processing a massive amount of sensory data collected by multiple sensors, which helps autonomous vehicles understand the surrounding environments and accordingly execute driving activities. In addition to CAVs, the emergence of flying platforms such as unmanned aerial vehicles (UAVs), commonly known as drones, enables several key potential 6G applications and services in a broad spectrum of domains thanks to their mobility, flexibility, and adaptive altitude.

2) REQUIREMENTS

Besides some key connectivity requirements to achieve high-speed, real-time, and reliable data transmission with different communication scenarios, including vehicle-to-vehicle (V2V), vehicle-to-infrastructure (V2I), vehicle-to-cloud (V2C), vehicle-to-pedestrian (V2P), and vehicle-to-everything (V2X), CAVs should strictly demand about the QoS performance of autonomous driving systems. The requirements may vary according to the autonomous level of vehicles: no driving automation, driving assistance, partial automation, conditional automation, high automation, and full automation. To achieve the baseline autonomous requirement, a vehicle needs to be aware of its surroundings during the driving period by first perceiving information and subsequently acting with vehicle control. To fully understand the driving environment, a massive amount of data recorded by multimodal sensors should be processed automatically and accurately via a driving computer system with advanced AI/ML models. Due to the critical demand for safety, a self-driving AI-powered system is expected to operate flawlessly regardless of weather conditions, visibility, road surface quality, and other situational conditions. In this context, advanced ML algorithms and DL architectures have been recently exploited to cover all possible driving scenarios and surrounding environments [185], however, the trustworthiness of AI systems embedded in CAVs is questionable. In addition to delivering useful driving recommendations and ensuring accurate driving activities promptly to minimize the probability of traffic accidents, a revolutionary AI system should be both understandable and explainable. This ensures that drivers feel confident about their decisions [186].

3) EXISTING AI SOLUTIONS

For safe and efficient operation on roads, CAVs should understand the current state of nearby vehicles and surroundings to predict future driving behaviors, which allows AI systems to react automatically and immediately. In [183], a comprehensive survey on AI-aided driving behavior prediction and potential risk analysis was presented, in which the advances of DL with different architectures, compared with conventional ML algorithms, were deliberated in terms of prediction performance. This survey showed that DL has been represented as a promising solution to deal with different sensors (e.g., cameras, LiDAR, and radar) for complicated driving scenarios. The survey also categorized the state-of-the-art driving behavior prediction approaches into three classes: input representation (track history of target and surrounding vehicles, bird's-eye view of the environment, and raw sensory data), output type (maneuver intention, unimodal trajectory, multimodal trajectory, and occupancy map), and prediction method (RNN, LSTM, and CNN). In [187], an innovative road condition supervision system was developed by learning a deep network to reduce accidents caused by poor road quality, in which the sensory data of the accelerometer and the gyroscope were processed in coordination with GPS data. Suddenly changing lanes and braking of the leading vehicle, caused by driver's distraction, misjudgment, and misoperation, increase the risk of accidents. An autonomous braking decision-making strategy was proposed with deep reinforcement learning in [188] to facilitate low-level control of CAVs in emergencies. Despite promising as the core of next-generation intelligent transportation systems, CAVs will face some hidden security problems relating to the complicated AI configurations/settings of autopilot systems [189], where end-users with limited AI experience and knowledge can be a nuisance and may cause accidents and risky dangers. ML and DL have been recently leveraged for various UAV-based applications in wireless communication systems (e.g., interference management, catching optimization, and resource allocation [190]) and intelligent transportation systems (e.g., trajectory planning, traffic flow monitoring, and navigation [191], [192], [193]). In cell-free and aerial-assisted vehicular networks [194], UAVs were used to assist CAVs in making decisions in driver assistance systems (for example, path planning and crash warning on highway), in which a supervised learning model was designed with CNN architectures and optimized with particle swarm optimization to make timely inference and facilitate online decisions in real-world scenarios.

4) HOW XAI CAN HELP

The important role of AI in driver assistance systems in CAVs is undeniable as we move forward to the next generation of intelligent transportation systems. However, there is an arguable issue of whether drivers completely feel confident and secure with AI-based decisions. Several advanced assistance systems developed by some big companies like Tesla are very complex and have numerous AI-based functionalities (e.g., automatic parking, adaptive cruise control, automotive navigation, collision avoidance system, driver drowsiness detection, and vehicle stability control), which ask end-users to tune multiple settings to ensure that the systems will operate smoothly and properly. The transparency of system operation and the explainability of decision-making are very immature in driverless cars. For example, a self-organizing neuro-fuzzy model coupled with a density-based feature selection technique was leveraged in [195] to explain automated reacting decisions (e.g., braking, speeding up, changing lane to left), in which the outputs of an AI model were capable of being interpreted by human-understandable if-then rules. In [196], the rule-based XAI was also investigated for different flying events of UAVs, in which the decision of changing the lying path can be explained regarding the weather conditions, surrounding environments, and relative enemy locations based on the representation of if-then rules derived from a fuzzy inference model. In another work [197], XAI with a random forest algorithm was proposed to improve the self-confidence level of autonomous navigation in an autopilot system. The intermediate information extracted from trees helped explain navigation autonomy and AI-driven decisions. Most of the existing works of XAI for CAVs have focused on explaining the decision of a single system instead of the decision derived from multiple interactive systems, where changeable environments and conditions can affect the outputs of prediction systems and consequently lead to some potential threats. Indeed, explaining why a vehicle changes a lane or hits the brakes is a more complicated task. For an early risk-aware system, a hierarchical AI model [198] was studied to predict uncertainty and collision risks under the constraint of account perception, intention recognition, and tracking error.

At the highest level of automation (i.e., a vehicle is free from geofencing with the capability of reacting as an experienced human driver), explainability is one of the most important requirements for not only system developers and

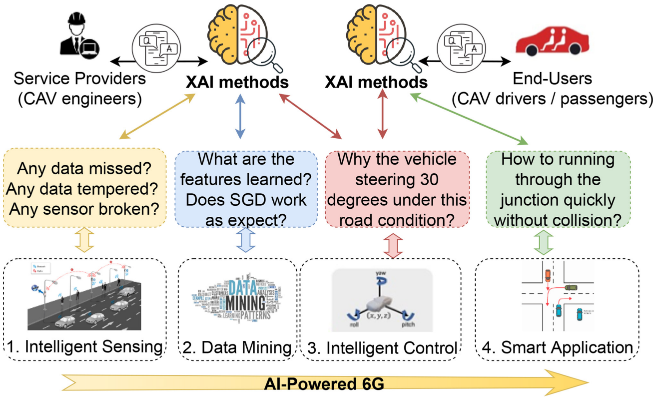


FIGURE 10. An illustration of typical questions that XAI can answer for various CAV stakeholders (e.g., engineers and passengers) across all AI-powered 6G layers. Note that the three first questions (left-right) are related to *Service Providers* and the two last questions are related to *End-Users*.

manufacturers but also customers, passengers, and society. Fig. 10 shows some exemplary questions CAV stakeholders may raise from each intelligent 6G layer that XAI can help to answer. Specifically, for the service providers, XAI will improve the quality of data that is collected in the 6G intelligent sensing layer across heterogeneous devices (e.g., vehicles, pedestrians, UAVs) and wireless network conditions (e.g., highway, congested/un-congested urban scenarios), ensure the trained AI models on the 6G data mining layer are robust and generalized to diverse scenarios (e.g., traffic predictions for major/minor roads), and improve the efficiency of network resource management in the 6G intelligent control layer in supporting all intelligent functions (e.g., more switches or routers allocated to congested urban scenarios.). Those (i.e., sensing layer, data mining layer, and intelligent control layer) are the main changes required in 6G when compared with existing 5G architecture. For the end-users, at the 6G smart application layer, XAI will provide more detail on each driving behavior decision to enhance the trust of drivers/passengers in CAVs.

D. SMART GRID 2.0

1) MOTIVATION

The core infrastructure in the smart grid is provided by IoT. The smart grid, characterized by automation, informatization, and interaction, provides a diverse and quality power supply for customers efficiently and securely. To provide supervision of the assets reliably, information interaction in real-time, peer-to-peer trading of energy, load management, and other electrical services, the smart grid needs a communication infrastructure, which is flexible and also highly reliable [199], [200]. One of the challenges in making the smart grid more sustainable is to manage the remote communication among various systems that are connected by smart meters [201]. Smart grid 2.0 is a futuristic evolution, where seamless connectivity of several power generation sources such as large-scale renewable energy sources is offered. In smart grid 2.0, machine-to-machine communication is facilitated through AI-based algorithms with no third-party intervention to automate the operations of the smart grid. Customers can choose economically viable local micrograds and at the same time pay attention to reducing the impact on the environment. Smart cities, next-generation vehicles such as electric vehicles, unmanned aerial vehicles, and autonomous vehicles would benefit from their capability of locating the nearest charging pile that offers the best electricity price [202], [203], [204]. In smart grid 2.0, smart autonomous contracts will autonomously execute all the exchanges of energy transactions if they meet the pre-defined conditions. In the futuristic smart grid 2.0 the real-time measurements are acquired through IoT sensors, which would be processed in distributed edge devices and the data would be transmitted to energy management systems that use cloud computing. The fusion of big data analytics and deep learning-based algorithms would help in realizing self-resilient smart grids [205]. 6G can help in realizing intelligent applications of smart grid 2.0-like remote monitoring and remote controlling of distributed energy resources, automation of demand response, etc. [206]. A smart meter needs the deployment of a distributed network, and wide coverage is essential in preventing blackouts and also to ensure the smart grid's self-healing capabilities. 6G can also help in realizing applications of smart grid 2.0 that require high-speed connectivity, such as predictive maintenance, video surveillance in real-time or during natural calamities, recovering proactively during emergency times, etc. [207], [208].

2) REQUIREMENTS

To deal with the massive volume of data generated due to constant communication and connectivity in the smart grid, sophisticated techniques that can analyze the data and assist in the decision-making process are required [209]. ML can solve the problems arising because of the large volumes of data generated from smart grids and can assist the smart grids in the collection of data, analyzing the patterns existing in the data, and also making decisions to run the smart grid. An ML-enabled 6G network can benefit the smart grid by solving some of the issues in real-time such as automated detection of intruders in the network, forecasting the price of electricity consumption, electricity thefts, line maintenance, generation of power based on demand, optimal scheduling, detection of faults, demand response, prediction of the stability of a smart grid, etc. [210].

3) EXISTING AI SOLUTIONS

To address the security of the smart grid, Babar et al. [211] proposed a secured demand-side management engine for the smart grid using Naive Bayes, a machine learning algorithm, to preserve the energy utilization in the smart grid. Due to the increased data generated at a rapid pace in 5G and beyond, enabling the smart grid, data acquisition, and processing by a smart meter is vital. The redundant data present in the data acquired can be reduced by using event-driven sampling. To address this issue, Qaisar et al. [212] employed the SVM

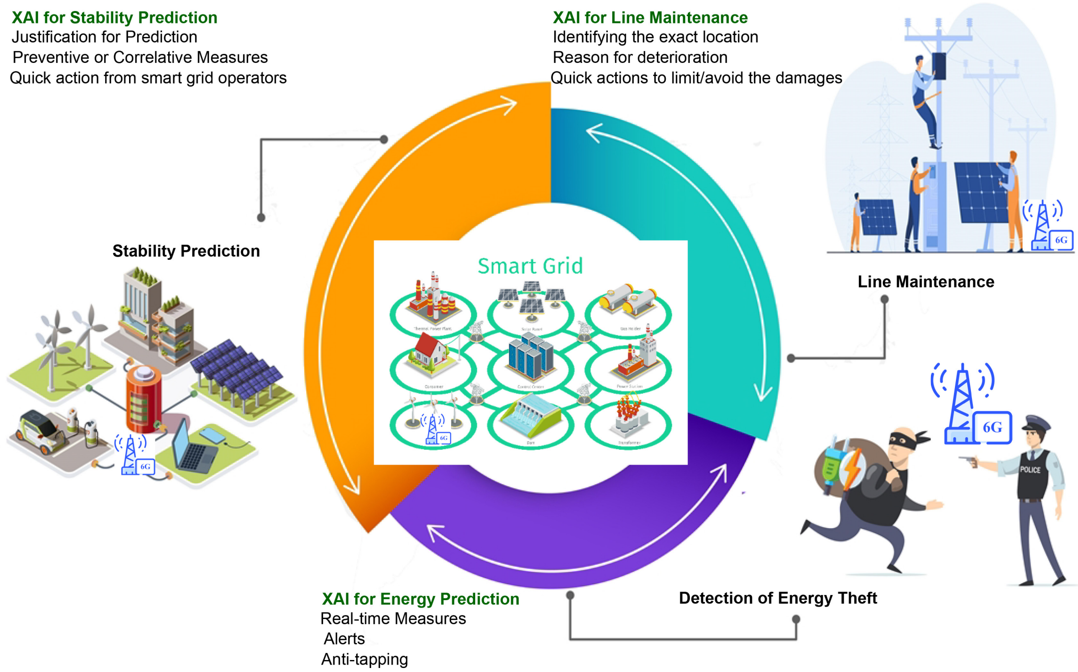


FIGURE 11. An illustration of XAI for 6G-enabled smart grid 2.0: stability prediction, energy theft prediction, line maintenance, and charging pile localization.

algorithm to identify the relevant features for analyzing the consumption patterns of appliances. Providing on-demand services for electric vehicles through vehicle-to-grid systems is very important because of the height maneuverability of electric vehicles. Shen et al. [213] proposed a hybrid architecture based on cloud and fog computing with applications in the 5G-enabled vehicle-to-grid networks. The proposed architecture allows the bi-directional flow of information and power between smart grids and schedulable electric vehicles to improve the cost-effectiveness and QoS of the energy service providers. For proper scheduling, the selection of suitable electric vehicles is very important. To improve scheduling efficiency, finding the categories of target electric vehicle users is required. To identify target electric vehicles, the authors proposed an artificial intelligence method that is based on the electric vehicle's charging behavior. In a similar work, Sun et al. [214] proposed a novel architecture for a 5G-enabled smart grid based on edge computing and network slicing for providing on-demand services for electric vehicles. This architecture collects the bidirectional information of traffic between the electric vehicles and the smart grids to decrease the cost of the energy providers and improve the charging experience of electric vehicles. To improve the scheduling efficiency of electric vehicles, the authors proposed LSTM-based electric charging behavior prediction, KNN-based classification of electric vehicle charging, and k-means-based clustering of electric vehicle charging.

4) HOW XAI CAN HELP

Using XAI in 6G-based smart grid systems will enable the collection of the most significant data through smart meters and sensing devices located at the intelligent sensing layer. This data will be subjected to XAI algorithms to generate justifiable/interpretable patterns in the data mining layer (demand response, fault line prediction, closest charging pile detection). Also, the use of XAI will ensure optimized resource management, efficient network automation using ZSM, and accurate information broadcasting through intelligent radio in the intelligent control layer of 6G architecture. These changes involving XAI in the sensing layer, data mining layer, and intelligent control layer are required in the 6G architecture when compared to the existing 5G architecture. In the application layer of the 6G architecture, the use of XAI will enable more informed and justifiable decision-making capabilities for the end users to ensure enhanced trust and confidence in the smart grid applications.

Kuzlu et al. [215] proposed a forecasting approach based on an XAI methodology to predict the generation of PV power, to increase the trustability of AI models, and hence improve the acceptance of AI in smart grid applications. Zhang et al. [216] proposed a Shapley additive explanations (SHAPs) based backpropagation deep explainer, termed as Deep-SHAP method, that produces an interpretable model for emergency control applications in smart grids.

Some of the use cases of XAI-enabled 6G for the smart grid are discussed below and are summarized in Fig. 11.

Use Case 1 (Stability Predictions of a Smart Grid): Maintaining the stability of a smart grid is of paramount importance. There are two criteria concerning the stability of a smart grid. The first criterion is to have a reserve of battery storage to meet the dynamic demand for electricity. The second criterion is providing enough capacity for the stability of the voltage at every location. The instability of a smart grid may lead to power outages and blackouts, which may lead to huge losses in revenue in several industries [217], [218]. ML algorithms can analyze the data from several sensors in the 6G-enabled smart grid to predict its stability. The AI/ML algorithms should be able to predict the stability of a smart grid early so that the operators can take the necessary actions [219]. Since traditional AI/ML algorithms could not justify the predictions, the operators of a smart grid might be reluctant to take preventive or corrective measures earlier to avoid losses due to instability. XAI can help smart grid operators identify the causes of the under-voltage load shedding, the need for secondary control power, frequency stability, deterministic frequency deviations, post-fault transient stability issues, and transient stability issues that make the smart grid unstable. The insights provided by XAI may encourage them to take immediate action without further investigation.

Use Case 2 (Detection of Energy Theft in Smart Grid): It is estimated that approximately 96 Billion dollars are being lost every year by the utilities due to energy thefts, which lead to increased prices of energy for the consumers [220]. The energy thieves make use of several methods such as tapping a line between a house and the transformer, hacking into meters of neighbors/their meters, and tampering with the meters [221]. To minimize electricity thefts, we have to identify the most likely cases of theft that can be investigated further. By training the ML models on the data from the smart meters, other external factors like geographic risk in a particular area, and weather in a 6G enabled smart grid, we can generate such a list in real-time that will enable the operators to take appropriate measures immediately [222]. There is no fixed solution regarding the action taken by the investigators/operators of the smart grid. For example, a meter that is reversed may have to be disconnected, if a meter is intruded, the household has to be alerted, and also the meter that is altered has to be replaced, etc. XAI can be helpful in this scenario to explain what kind of theft may happen so that the operators of the smart grid can take relevant action to address the issue. XAI can help smart grid operators can also identify energy theft caused by physical obstruction, electrostatic attacks on electronic meters, obfuscation of the energy meter, the introduction of foreign material into the energy meter, and direct line connection.

Use Case 3 (Line maintenance in Smart Grid): The reliability of a smart grid depends on the proper maintenance of the infrastructure. The deterioration/aging of a transformer or power lines (e.g., due to weather conditions) has to be detected at an early stage to prevent the failure of equipment that may cause power outages and blackouts [223]. The conventional AI/ML methods in a 6G-enabled smart grid can predict the aging/deterioration of the lines based on data generated from the sensors such as humidity, weather, and so on in real time. XAI can be used efficiently in these scenarios to identify the exact location and reason for the deterioration of the power lines so that the maintenance crew can be sent to the exact location with the heads up on the reasons so that they can take actions to limit/avoid the damages.

Use Case 4 (Location of Closest Charging Pile by Electric Vehicles): In the futuristic Smart Grid 2.0, electric vehicles can choose the closest charging pile that offers the best price for charging electricity. To realize this, a 6G communication network, which offers ultra-reliable and low-latency continuous services, is essential to monitor constantly the geographical location of the electric vehicles and also to keep searching for charging stations near the location of electric vehicles. ML algorithms can help to identify the requirement of the amount the charging for the vehicles and also the apt charging station. However, the black-box nature of the ML algorithms makes it difficult for electric vehicles to choose the charging station among the available piles. XAI can come in handy in this situation. With justification provided by XAI, owners of electric vehicles can get a clear picture of which charging stations can offer electricity charging at optimal pricing considering many factors such as the distance of the charging station from the location of the vehicle, incentives offered, etc. This will help owners of electric vehicles select a charging station based on the analysis provided by XAI. Hence, XAI-enabled 6G can play a vital role in realizing several futuristic and autonomous services of Smart Grid 2.0 for electric vehicles.

It is assumed that the XAI algorithms such as PIRL, SHAP, and Facets can be integrated with 5G and beyond cellular networks so that the grid operators can benefit from the justifications/explanations provided by the XAI algorithms that can help them in making apt decisions in several smart grid 2.0 applications/scenarios. Even though XAI has a tremendous potential to improve the operations of smart grid 2.0 by bringing in more transparency and justification of the results obtained by AI algorithms, some challenges are to be addressed to reap the benefits of XAI-enabled 5G and beyond in smart grid 2.0 applications. One challenge is that due to the lack of metrics to measure the performance of the XAI algorithms, grid operators may face a severe challenge while making decisions in real-time.

E. MULTI-SENSORY XR APPLICATIONS, HOLOGRAPHIC TELEPRESENCE, METAVERSE

1) MOTIVATION

Extended Reality (XR) is a combination of all immersive technologies such as virtual reality (VR), augmented reality (AR), mixed reality (MR), etc. These immersive technologies are used to extend the reality that is experienced by the creation of an experience that is fully immersive or by amalgamating the real and virtual worlds using multiple sensors [224], [225]. XR has many real-time and practical applications in many sectors where the travel cost and time for the customers can be saved, such as entertainment, retail, healthcare, real estate, marketing, remote working, disaster handling, etc. [226], [227].

Holographic Telepresence is a technology, where the systems can project real-time, full-motion, realistic 3D images of people located in distant places into a room, with real-time audio communication, which will make the users feel as though they are communicating with people in person. Unlike in AR or VR, users don't require any device, sensors, or headsets to experience holographic telepresence. In holographic telepresence, the captured images of people at remote locations along with their surrounding objects will be compressed and then transmitted through a broadband network. These images will be decompressed at the users' end and then projected through laser beams. Holographic Telepresence has the potential to revolutionize traditional communication via mobile phones by giving immersive experiences to users. It has huge potential in many other applications of communications such as telemedicine, enhanced television and movie experiences, gaming, advertising, robot control, aerospace navigation, 3D mapping, and other simulations [228], [229].

Metaverse is a trending term that is the convergence of two principal ideas: virtual reality and digital second life, and has been very recently attracting much more interest from various academic communities and big tech companies [230], [231]. Different from AR which can deliver to users the experiences of video streams and holograms in the real physical world, VR is responsible for conveying immersive experiences in the virtual world. With a VR headset, users can experience numerous services and applications in the metaverse, and create their hyperreal content [232], [233]. In this context, a variety of AI algorithms have been applied and developed in VR devices to improve the human-machine interactive experience based on modeling and learning visual information. Some metaverse projects (for example, Decentraland and Sandbox) have successfully built virtual reality platforms, in which AI contributes in many aspects, such as image retrieval, image quality enhancement, and 3-D video rendering.

2) REQUIREMENTS

XR and holographic telepresence technologies require a communication network that has near-zero latency, and fast processing of the information from sensors. 6G, through its attributes like connection density, user-experienced data rate, scalability, mobility, reliability, and traffic volume density, can play an efficient role in realizing the true benefits of XR [141], [234], [235], [236].

3) EXISTING AI SOLUTIONS

AI can be effectively used for self-managing the participating devices in 6G-enabled multisensory XR and holographic telepresence technologies. Some of the potential applications of AI in the end devices in these applications are to understand the environment by applying computer vision to analyze the multidimensional knowledge from the images captured by the devices, reduction of network volume by enabling AI-based applications in mobile devices, etc. Some of the potential applications of AI/ML in multi-sensory XR applications and Holographic Telepresence are [237]:

- *Estimation of the Position of Objects:* The object's position (such as fingers, hands) can be inferred for controlling the content of XR [238].
- *Labelling of Scenes and Images:* Triggering XR labels with image classification [239].
- *Semantic Segmentation and Occlusion:* Segmentation and occlusion of the specified objects [240].
- *Object Detection:* The object's extent and position in a scene can be estimated to form colliders and hitboxes to enable interactions between virtual and physical objects [241].
- *Recognition of Audio:* Triggering the effects of AR through recognition of keywords [242].
- *Recognition and Translation of Text:* The application interfaces of XR can be used for overlaying the text detected from an image into the 3D world [243].
- *Content Generation:* Designing of environment, characters, and other graphical objects [244].
- *Virtual Humans:* Training of animations so that they can respond in real-time [245].
- *Virtual Assistants for Dynamic Customer Experiences:* Training of virtual assistants that answer the queries of customers to provide a virtual experience on the latest trends [246].

4) HOW XAI CAN HELP

XAI can offer valuable reasoning and justification for predictions/classifications, persuading XR/Holographic Telepresence application providers to base decisions on AI/ML outcomes. For example, precise estimation of human hand/finger positions is crucial in controlling XR content. AI/ML algorithms can be used to estimate the position of the objects. Real-time decisions based on AI/ML recommendations may occasionally result in inaccurate XR content due to false positives. If the 6G is enabled with XAI, the application providers will understand the reasoning behind the predictions/classifications by the AI/ML algorithms that can help them make accurate decisions in real time. Similarly, if 6G is enabled with XAI, the virtual assistants can provide accurate information to customers [247].

In holographic telepresence-based applications, the use of XAI will enable the generation of realistic 3D images (e.g., human anatomy, clinicians, human holograph images) of individuals located in geographically distributed locations in the intelligent sensing layer. This data will be processed using XAI algorithms providing a realistic and immersive experience for the users without the use of any physical devices, or sensors for achieving optimum quality holographic presence. In the intelligent control layer, the use of XAI would enable seamless and faster processing of data via network automation and ZSM, optimized resource management ensuring near zero latency while streaming of the 3D images ensuring justifiable decision making by the stakeholders in the intelligent control layer within the 6G architecture. The aforementioned changes are essential for the integration of XAI into 5G architecture for realizing the true potential of holographic telepresence applications in 6G. The use of XAI will enable the realistic presence of the end users and the accurate positioning of objects in a holographic telepresence environment ensuring accurate decision-making at the 6G smart application layer.

The authors in [248] proposed guidelines for using XAI techniques and simulations using XR for secured human-robot interactions. The authors suggested that the proliferation of high-fidelity VR-based simulation environments will result in the reduction of barriers in cataloging and performing postmortems in operations by robotics that may result in the characterization of more rigorous behavior of autonomous system behavior and promote the adaption of explainable techniques in their controllers.

When XAI is integrated with 5G and beyond cellular networks, it may lead to performance degradation of the AI algorithms to enable explainability. In the metaverse, many VR-based services and applications leverage advanced AI models to enhance user experiences with interactive activities; however, they are usually presented as black boxes without interpretability and explainability. In the effort to completely explain AI decision-making processes, a variety of XAI algorithms and methods can be studied for many development tasks (e.g., object detection, semantic segmentation, image super-resolution, 3D video rendering, etc.) in the metaverse framework. For instance, with LIME and LRP, 3D designers and computer vision scientists who apply DL to build virtual worlds can understand and explain what is happening inside deep models and when they are likely to be broken down.

F. SMART GOVERNANCE

1) MOTIVATION

Smart governance is perceived as the intelligent use of ICT and innovation to facilitate and support enhanced decision-making, planning, and citizens' role through collaborative decision-making [249]. The motivation of smart governance is similar to the ones realized under the ideals of good governance [250] in modern-day democracies, with an additional focus on ICT to uphold the ideals, ensuring the development and welfare of the public and public resources [251]. The fundamental challenge that remains relevant in the existing governance is that of corruption [252] and unfair policies, and methods, to improve education, security, transport, resource management, and economic infrastructure, which is where smart governance is envisioned to offer better solutions.

Presently, 5G is enabling hyper-connectivity, decreasing latency, increasing traffic capacity, and improving throughput compared to 4G and its predecessor networks. Causing an evolution in smart governance applications due to unprecedented levels of real-time information from any device, anytime, and anywhere, while improving public infrastructure and experience [253]. The main beneficiaries of smart governance are the general public and the government. The demand for smart governance will grow with the growth of 6G to drive innovative applications enhancing the user experience of governance by collecting user data and rewarding with enriched information through the advancement in AI and XAI [254], such as finding the quickest path to a destination, following election campaigns, law enforcement decisions or guidelines, or postal service tracking.

2) REQUIREMENTS

To fully realize the vision of smart governance, we need ICT services with high bandwidth connectivity and more devices to connect and communicate with more focus on understanding decision-making. The understanding of decision-making means the Internet supports XAI, empowering users with the decision and accompanying explanations to keep the user informed. This decision-making and explanation would require real-time operations, such as while driving in cooperative traffic management; therefore, high latency would be catastrophic. Complex traffic management would require big data operation for faster and safer commutes when many devices and sensors produce data simultaneously. Besides this, by the time 6G goes into effect, the advancement in XAI will be further advanced, meaning smart governance applications would require support from infrastructure to consume XAI to its futuristic potential. It will empower users with transparency, meaning 6G should allow application support in terms of explanation instead of focusing on improving hardware and signal processing.

Overall, the idea of the Internet of Everything will be a crucial requirement for smart governance, requiring all sensors and devices spread across smart cities to achieve high-speed and real-time connectivity. Also, data transmission would demand highly reliable data exchange to ensure QoS performance and optimal service to citizens, the government, and other stakeholders. Additionally, to implement 6G smart governance, the key difference compared with the existing 5G network is that a much higher level of security and privacy is required, especially at the intelligent sensing layer of the 6G architecture shown in Figure 2. It is crucial to ensure the trustworthiness of the massive data collected from diverse sensors by the network itself, as humans cannot inspect data on such a large scale. Moreover, as most of the data collected in smart governance scenarios may endanger the privacy concerns of personal data, the whole data collection process needs legal compliance, which XAI has great potential to ensure through transparency.

3) HOW XAI CAN HELP

The 5G initiative has primarily focused on enhancing network infrastructure, sometimes overlooking the application side of services. Hence, it restricted applications from fully exploiting XAI in the context of smart governance. Such as the current smart monitoring services have limited use of transparency in decision-making [255], which is one of the reasons engineers don't fully trust automated decision-making, where human oversight plays a crucial role in monitoring. For example, in the oil and gas sector, monitoring and maintenance of Valves are critical for maintaining steady operations. Here the AI helps with decision-making by informing when a valve is degrading and needs to be replaced; however, the AI does not explain the decision, which makes it hard for hardware engineers to trust the decision. Things can go wrong when a valve is coming to the end of life, and the AI decides it is healthy. From a governance point of view, it will be highly catastrophic if oil gets leaked into the sea, causing significant environmental harm. Therefore, to avoid this circumstance, AI needs to explain the decision to bring transparency, empowering hardware engineers to make an informed judgment. To this end, recently some efforts have been made to explain decisions through graphs and reports based on semantics that allow engineers and decision-makers to understand automated decisions better [256], [257]. However, the efforts could be advanced using XAI approaches such as LRP, Bayesian RL, LIME, and SHAP that can improve the quality of explanation by embedding explanation as a part of the system instead of as a mere add-on; this will be strengthened through 6G transparency and high-speed network. Also, here, explanations can indicate what has changed visually in the video feed or through sensor data points that may call human attention for monitoring purposes. Another area of application is public engagement in the system of governance [258]. At present, governments are opting for limited use of dashboards when it comes to smart cities, and smart governance [259] to share insight and stats. In this regard, there were some applications developed to bring insights during the election campaigns [260] and movements of public concerns and social causes [261] (e.g., austerity, Brexit, refugee crisis) with the use of AI. These types of dashboards bring some transparency, such as which politician or political party has a specific stance concerning a domestic political issue on their social media, what is the public voice concerning austerity, and the voice of different media outlets. However, these dashboards don't explain their decisions to the degree where XAI is making advancements. With XAI's inherent strength of explaining the decision and 6G's interface that ensures high-quality real-time data ingestion, these dashboards will further advance and play essential roles during the elections, referendum, and other political processes for all political stakeholders (public, politicians, and government). In these dashboards, XAI approaches will further open doors for explaining the insights, accompanied by visualizations that portray real-time data, particularly during elections. These explainable insights can take the form of visuals, text, or even be conveyed through conversation, where the likes of surrogate, visual, and textual XAI approaches will contribute to improving the overall user experience in explaining insights.

Different XAI algorithms can be applied in the context of smart governance, where visuals and reports based on semantics will be beneficial [256]. Model-agnostic approaches [262] can be beneficial for complex decision-making algorithms (such as DNN) where explanations may be demanded in a real-time setting (applications such as route planning and smart monitoring). However, model-specific approaches will be useful where the precision of the decision's explanation is significantly essential.

These types of decision-making may not require real-time processing, such as when legal auditors are among stakeholders or when decision-making informs policy-makers and other applications can be city infrastructure planning. Both model-agnostic and model-specific can also co-exist for several applications within smart governance. Here, model-agnostic approaches can provide a simpler and quicker explanation that may be useful for end-users. For a more detailed and accurate explanation, model-specific explanations can be provided. The applications here would include all areas of smart governance, such as smart monitoring, election campaigns, causes of public concern, law enforcement, and public order.

Despite the great potential of XAI in smart governance, a simple explanation through a simplified user interface of visuals or reports-based explanation might not inform the user of an accurate explanation. It may mislead individuals and become critical when applications belong to causes of public concern. Here, application providers must inform users of the potential risk of simplified explanations. Also, there is a danger of trying to explain election campaigns due to a lack of metrics of explanation, as the campaigns might be so dynamic and novel that it would be challenging even for a human expert to explain them due to subjectivity, let alone XAI to do the job.

However, acknowledging the shortcomings of technology may help. On the contrary, a complex explanation may be more accurate but hard to comprehend. Also, real-time decision-making and explaining the decision may be critical for law and order applications requiring accurate information that might demand computational power beyond current capabilities, which may be achieved through technological advancements. In particular, for stakeholders, XAI will offer facilities to citizens and help better govern cities and infrastructure, which is closely linked to data quality and reliability at a high volume, a facet ensured by 6G. Also, until IoE is fully realized, the promises of smart governance, cities, and offered smart infrastructure cannot be truly achieved [267]. Especially most of the network support for IoE would be wireless, which is where promises of 6G are important. From a stakeholder's perspective, getting informed about a high volume of data from IoE means a correct decision that can be trusted through explanation. Overall, for smart governance, the efficiency of mobility and resource management are critical issues that would be potentially dealt with through the vision of 6G, similar to Open RAN real-time operations via xAPPS deployed in NR-RIC.

TABLE 5. The requirement analysis of XAI for typical B5G/6G use cases.

6G Use Cases	Typical High-Stake AI Decisions	Key Stakeholders[a]			XAI Challenges[b]			
		Service Providers	Legal Auditors	End Users	Explainability Metrics	Exp. Perf. Trade-off	Legal Engagement	Privacy Leakage
Intelligent Health and Wearable, Body Area Networks	Medical Image Diagnosis, Health Risk Prediction, Treatment Planning and Supervision	M	H	M	M	L	H	H
Industry 5.0, Collaborative Robots, Digital Twin	Predictive Maintenance and Analytic, Quality Inspection, Machine Condition Monitoring and Supervision	H	M	M	H	H	L	L
CAVs and UAVs	Collision Avoidance	H	H	H	M	H	H	M
Smart Grid 2.0	Predictive maintenance, Theft detection	H	M	L	M	M	M	L
Multi-sensory XR Applications, Holographic Telepresence, Metaverse	Remote operation of machines	H	M	H	M	L	M	H
Smart Governance	Policy recommendation	M	H	H	H	M	H	H

[a] Detailed discussions can be found in Table 3.
[b] Detailed discussions can be found in Section VI.

L Low Demand M Medium Demand H High Demand

G. SUMMARY OF THE XAI IMPACT ON 6G APPLICATIONS AND TECHNICAL ASPECTS

This section reviews the attempts and potentials of XAI for a wide range of 6G applications. Different from 6G technical aspects, 6G applications involve more types of stakeholders than engineers only: such as end-users and legal auditors. Therefore, the requirements of XAI for 6G applications will have to be analyzed case by case. In Table 5, for each 6G use case, we describe its typical high-stake AI-powered decisions that need the XAI most. Then, we identify the level of demand for XAI for each stakeholder, and for each XAI challenge that needs to be addressed in the future. For instance, collision avoidance of CAVs or UAVs is a typical high-stake AI decision as the incorrect decision can lead to the loss of life. Thus, all stakeholders would need the explainability for such decisions at the highest level. Moreover, collision avoidance requires both high model accuracy and explainability to give evidence for legal experts to judge responsibilities on various occasions. Another example from Table 5 is the quality inspection of Industry 5.0. If a product is mistakenly qualified, it would be risky for service providers mainly as it affects their reputation. As most Industry 5.0 activities are within the factory, the demand for wider legal engagement and higher privacy protection is low.

In Table 6, we also summarised the key references that are discussed in both this section and the previous section about 6G applications and technical aspects. Lots of existing work focused on the security, privacy, and edge AI technical aspects as these contain the most high-stake decision-making process. In addition, these technical aspects also have AI solutions deployed already in many existing systems. However, the lower-level intelligent radio and the backhaul ZSM are lacking attention, partly because these technologies are still being studied in the early stages. The existing research interests are roughly evenly distributed across all 6G applications discussed in this paper. CAVs and UAVs are the ones that have attracted the most interest so far due to their large existing research communities, while intelligent health still demands more work as it requires close interdisciplinary collaborations.

VI. LIMITATIONS AND CHALLENGES OF XAI FOR 6G

This section summarises the major limitations of the recent research and the possibilities of applying XAI for 6G. The section also provides the research challenges to moving XAI for 6G forward.

A. LIMITATIONS OF XAI FOR 6G

There are several well-known limitations [25] of existing XAI methods that would also delay the successful deployment of XAI in the future 6G infrastructure.

- *Lack of in-model XAI methods:* There is a widespread concern [19] that the performance of the AI models will go down with the growth of its corresponding level of explainability because of the ever-increasing level of AI model complexity. Recently, in-model XAI methods are likely to be satisfactory in both AI performance and its corresponding explainability. It is because the in-model XAI methods are designed to be self-explanatory, rather than an add-on to the XAI method after the AI decision is made (i.e., post-hoc XAI methods). However, most existing XAI methods are still post-hoc (e.g., LIME, SHAP), as they are more straightforward [25] to be developed and pluggable with existing AI algorithms compared to in-model methods.

TABLE 6. Summary of key related works in XAI algorithms and goals for 6G use cases and technical aspects.

Ref.	XAI Algorithm	Why XAI	Explanation	6G Use Cases						6G Technical Aspects						
				Intelligent Health	Industry 5.0	CAVs and UAVs	Smart Grid 2.0	XR, Holographic Telepresence	Smart Governance	Intelligent Radio	Trust and Security	Privacy	Resource Management	Edge Network / AI	Network Automation, ZSM	Other technical aspects
[263]	6-layer Decision Tree	Enhance the trust management for the Intrusion Detection System.	Rules		✓		✓				✓				✓	
[264]	LIME	Analyzing the rationale behind the unsupervised machine learning model for encrypted video traffic clustering.	Feature Importance					✓								✓
[265]	SHAP	Analyzing various machine learning models when predicting the bit error rate of the light path.	Feature Importance		✓											✓
[142]	SHAP	Demystify the behavior of complex AI models used in short-term resource reservation within sliced networks.	Feature Importance										✓		✓	
[264]	LIME, SHAP, PI, Eli5	Evaluate the performance of multiple XAI methods in identifying the cause of SLA violation prediction.	Feature Importance								✓				✓	
[45]	LIME, SHAP	Achieve a deeper understanding of ML algorithm behaviors in automated failure-cause identification in microwave networks.	Feature Importance		✓					✓	✓					
[162]	Grad-CAM	Improving DL explanations of prostate lesion localization while keeping high accuracy of lesion classification.	Visual	✓							✓					
[177]	Grad-CAM	Analyzing the rationale behind the AI model CNN in bearing faults diagnosis by analyzing the vibration signals.	Visual		✓									✓		
[181]	Bayesian inference	Generated explanations significantly improve the performance and user perception of the human-robot collaboration.	Causal		✓						✓					
[196]	Rule-based XAI	Explain the lying path decision of UAVs regarding weather conditions, surrounding environments, and enemy locations.	Rules			✓				✓				✓		
[215]	LIME, SHAP, Eli5	Gaining insights from AI models in predicting solar PV power generation for smart grid.	Feature Importance				✓									✓
[108]	LIME	Using fewer data to achieve better performance of GAN-based privacy-preserving solution by XAI's explanations on the discriminator's decision.	Feature Importance								✓	✓				
[261]	SHAP	Using XAI explanation to improve the feature engineering for earthquake-induced building-damage mapping.	Feature Importance						✓							✓

- *Lack of quantifiable explainability metrics:* Visual and textual explanations are two commonly observed formats of XAI methods output. These explanations are intuitive for human beings but difficult to measure objectively using quantifiable metrics. Therefore, it is challenging for XAI system designers to achieve standard/unified systems that are simple to deploy and use for all stakeholders.
- *Lack of engagement of stakeholders and legal experts:* A strong motivation for introducing XAI technologies is to address the legal requirements. A typical example is the "right to explanation" in the EU, GDPR which requires machine algorithms to be capable of giving explanations for their outputs. In the last few years of early research into XAI, computer scientists have proposed and applied many new technologies. However, two important points are missing. Firstly, a meaningful engagement of legal experts is required to ensure that XAI complies with legal requirements. Secondly, a deep engagement of stakeholders is needed to ensure the explanations provided by the new XAI methods make good sense to them.
- *Lack of privacy leakage consideration:* XAI could also increase the risk of privacy leakage [17] which is more significant in the 6G environment due to pervasive connectivities. This possible privacy leakage refers to the fact that when XAI is applied, more information will inevitably be exposed externally concerning the AI decision-making process, which likely leads to the leakage of users' data. Anonymization might be a possible solution to protect private information. However, if one can easily violate such protection, the risk of privacy leakage is still high.

B. RESEARCH CHALLENGES OF XAI FOR 6G

To address the identified XAI limitations for 6G, we discuss their respective research challenges in the following subsections.

1) DEVISING 6G-COMPLIANT QUANTIFIABLE METRICS TO ASSESS THE EFFECTIVENESS OF EXPLAINABILITY

When the DARPA XAI program [19] launched in 2017, researchers were focused on proposing general assessment frameworks or metrics across different domains. Doshi-Velez and Kim [268] proposed a taxonomy of XAI assessment methodology which contains three classes: application-grounded (i.e., measured for specific applications), human-grounded (i.e., measured for specific stakeholders), and functionality-grounded (i.e., measured for specific AI algorithms). Hoffman et al. [269] discussed the evaluation of XAI in depth from both psychometric and AI perspectives. They proposed an evaluation process for measuring the goodness of explanations, user satisfaction, user understanding of the AI system, user motivation for explanations, user trust and reliance on the AI, and the performance of the human-XAI work system.

TABLE 7. Key recent attempts at measuring the explainability of the XAI outcomes.

Ref.	Year	Metrics Proposed	Quantifiable	Domain Specific	6G Applicability
[268]	2017	Application-grounded, human-grounded, functionality-grounded	No	General	Medium
[269]	2018	The goodness of explanations, user's motivation for explanations, the performance of human-XAI work system	No	General	Medium
[270]	202	SCS	Yes	General	Medium
[23]	2020	QoT	Yes	6G	High
[15]	2021	Trustworthy Explainability Acceptance	Yes	Health	Low

Holzinger et al. [270] proposed a system causability scale (SCS) inspired by the success of the system usability scale (SUS) that has been widely used for assessing the usability of the system-human interface for over three decades. The SCS has ten questions to measure if XAI-generated explanations quickly meet the users' intention. In 2022, researchers from the Shapash open-source community[3] proposed metrics to assess the quality of explanations by XAI methods [271]. For example, *stability* measures for a certain XAI method if its generated explanations are similar for similar input data. *Consistency* tells of the same set of input data, and how explanations vary among different XAI methods. *Compacity* shows that a given XAI method can explain the majority of its input data by the minimum number of features.

XAI researchers soon realized that to make explainability measurement more effective, the quantifiable metrics have to be designed specifically for stakeholders and scenarios and cannot be domain-agnostic. For example, in the AI health domain, Kaur et al. [15] proposed a metric called "Trustworthy Explainability Acceptance" that measures the Euclidean distance between XAI explanations and domain experts' reasonings in predicting Ductal Carcinoma in Situ (DCIS) recurrence using AI. In the computer network domain, Li et al. [23] proposed a metric called quality of trust (QoT) to quantify the level of trust when a particular XAI model is applied for 6G applications. The QoT contains a physical and emotional trust, representing the objective and subjective assessment of explainability, respectively. The key attempts of the research mentioned in this section are summarised in Table 7.

The Quality of Trust (QoT) [23] serves as an exemplary initial effort in proposing quantifiable XAI metrics for 6G.

[3]https://github.com/MAIF/shapash

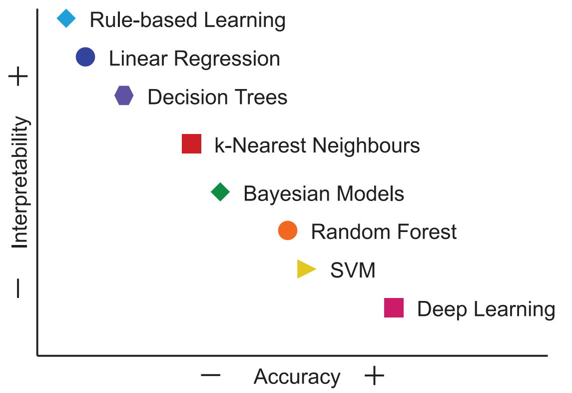


FIGURE 12. Representation of the Interpretability versus Accuracy according to [273]. Because of the increasing model complexity (i.e., from linear to non-linear) and training data size (i.e., both in volume and number of features), it normally shows the trend that the model interpretability decreases with the increasing model accuracy.

However, there are still many challenges to implementing a full set of XAI metrics for all key stakeholders in different 6G scenarios. For example, as AI will be used across all layers of the computer network infrastructure, the number of specific 6G AI scenarios that require explanations will be significantly higher than the existing 5G networks. Do we need to propose a set of metrics for each of these 6G scenarios? Can we reuse some metrics across various 6G AI scenarios? What are the metrics that we have to design specifically for particular use cases? Additionally, although SCS [270] is a good strategy for measuring user satisfaction with the generated explanation, the low-latency 6G networks may need quick solutions to capture the users' feedback. Moreover, for use cases where the stakeholders are engineers or scientists such as radio spectrum allocation, it is necessary to gain expert explanations in advance to ensure high effectiveness of XAI explanations, for example, [15]. However, the acquirement of explanations from domain experts needs to be planned carefully to avoid the huge potential cost in time and labor. The future research could also explore information theory approach in measuring the information gain before and after the application of XAI approach to a specific problem.

2) PROPOSING NEW XAI METHODS THAT CAN ACHIEVE A BETTER TRADE-OFF BETWEEN INTERPRETABILITY AND MODEL PERFORMANCE IN LARGE-SCALE 6G INFRASTRUCTURE

In the past decades, researchers primarily emphasized AI performance, with limited attention to interpretability [272]. However, the GDPR shifted the focus, a regulation recognizing "an explanation of the decision reached after assessment" for the users and holding automated algorithms accountable. The caused shift is now deriving from the trend towards the interpretability of the models.

The trade-off between the performance (accuracy) and the simplicity (interpretability) of a model has been studied many times in the literature [273], [274], [275]. As seen in Figure 12, the more complex a machine learning model is (such as a higher number of nodes, more rules, branches, or layers), the less likely it is to be interpretable. Adding complexity to the model is likely to model complex decision boundaries, making the model prediction more accurate. The fundamental challenge is to ensure higher accuracy without compromising on interpretability. In some domains, interpretable methods can provide similar levels of accuracy, and therefore they are recommended [276]. Hence, when choosing algorithms for constructing the large-scale 6G infrastructure, it's crucial to understand whether accuracy or interpretability holds more significance in a given domain.

Choosing an appropriate method that balances interpretability and performance is crucial, and the decision should depend on the specific domain and application. In certain fields, such as routing IP packets over a network, understanding the internal reasoning of the model may not be essential, as long as it demonstrates high efficiency. Conversely, in domains like medicine or critical decision-making within organizations, comprehending the model's internal reasoning becomes paramount. Alternatively, some perspectives recommend avoiding black-box models altogether due to their tendency to obscure the inference process, posing an increased risk of errors without a clear understanding of the underlying causes. Similarly, deep learning methods that automate feature selection prevent developers from identifying important features, as they often mix them with redundant ones [276], [277]. This significantly hampers the ability to overcome black-box challenges. To address the limitations of black-box models, there are two approaches. Firstly, advocating for the use of simple models with limited but acceptable performance has been proposed [277]. Secondly, the DL community is actively working on developing improved and more effective explainability techniques, which represents a growing trend [278].

Making informed decisions about algorithmic design is crucial for each module in 6G. While some modules, such as efficient IP packet routing or antenna selection for users, prioritize accuracy over interpretability, others, like providing route suggestions for cars, require models that are interpretable to a greater extent [5]. Finally, to strike a balance between AI model performance and interpretability, the open-source community has shown a clear inclination towards tree-based black-box AI models, such as random forest, LightGBM [279], and XGBoost [280]. These models, which require fewer data compared to DNNs, can still achieve high performance. Furthermore, reliable XAI methods, such as SHAP, offer specialized versions like TreeSHAP [281] for tree-based models. These versions provide relatively fast and reliable interpretability.

3) IMPROVING SOCIETAL AND ECONOMIC ENGAGEMENTS IN INTEGRATING XAI WITH 6G INFRASTRUCTURE

The third challenge of adopting XAI for 6G is about societal and economic engagement. Specifically, we first examine trends concerning laws and ethics. Then, we discuss commercial challenges concerning technology producers and intellectual property. Finally, we present the need for new laws and regulations to evolve and support XAI for 6G.

- *Laws & Ethics:* The rising demand and development of XAI and IoT will continue to shape evolving challenges in ethics, granting users more choices and preferences for transparency in automated decision-making. Moreover, with the growth of IoT to IoE within 6G, more devices would be connected with decision-making capability, raising security and privacy concerns. Several devices will probably interact directly without requiring human intervention. However, the endpoint will serve human demands. Here, additional laws and an understanding of ethics would be in demand to ensure humans control all activities. Laws like GDPR are necessary to mention, which will lead and enable innovation with the optimal level of governance over automated decision-making. However, GDPR alone does not encompass a full spectrum of sets of rules all around the globe. In addition, the Chinese PIPL and other regional laws would collectively undertake the future challenges of ethical practices concerning automated decision algorithms and devices.
- *Commercial Side:* Requiring decision-making algorithms to be explainable through legal or practical means can raise concerns for technology producers. The producer would find making algorithms transparent a risky business to protect the intellectual property of the technology. This concern would further grow when 6G starts to converge on the application-centric approach to automated decision-making, with the vision of IoE. It is here, that the approach taken by the U.S. is vital to consider which views data protection and data integrity as a commercial asset, unlike GDPR. It is important to note that explaining decision-making can leak critical information to competitors, which can quickly compromise commercial assets and undermine freedom of ownership. The right approach would balance GDPR and U.S. laws regarding data privacy and protection while safeguarding consumer rights and businesses.
- *Compliance of New Laws:* With the boom of 6G, continuous evolution would take place from IoT to IoE, focusing on applications that could explain automated decision-making. However, unlike hardware infrastructure, software applications get upgraded quickly, and many new types of applications keep popping up. However, at this pace, future applications necessitating explanations for decisions might face limitations due to stringent laws lacking the flexibility to accommodate various emerging requirements. Here, the laws should tolerate flexibility to assert the law's spirit that encourages best practices. These laws should protect consumer rights and commercial assets, ensuring both personal freedom and freedom of ownership. Finally, future laws should promote globally accepted rules while guaranteeing regional and transnational freedoms that provide some adjustment and flexibility. Otherwise, international acceptance of law that confirms the right to explanation to exact detail internationally may be too ideal to agree upon before 2030. The common future data privacy laws should strike a balance that protects businesses and users from exploitation while respecting national policies globally.

A solution for developing current laws with a balanced approach that recognizes consumer rights in terms of ethical considerations and ensures technology producers' intellectual property concerns. In addition, these laws should be flexible to allow rapid application development that matches the pace of 6G adoption. Since the upcoming 6G is a global phenomenon, future products would benefit from common international law. This common international law would better connect the world, world, safeguarding consumers' rights and addressing technology producers' concerns while promoting the rapid growth and international adoption of 6G products.

4) CONSIDERING THE PRIVACY AND SECURITY IMPACT WHEN USING XAI

The utilization of XAI raises concerns regarding privacy and security, particularly in 6G networks. XAI's explanations can play a vital role in detecting potential privacy violations associated with 6G. For instance, Slack et al. [14] demonstrated that explanations generated by LIME and SHAP, which remain uncompromised, can expose if an AI model's decision heavily relies on sensitive personal information such as gender, ethnicity, race, and others, which has implications for applications like credit scoring and the prediction of recidivism risks within the domain of 6G smart governance.

However, it is important to note that these explanations reveal additional information about the AI decision-making process and its underlying training data [282]. Consequently, XAI can inadvertently facilitate the transition of a black box attack into a white box attack, potentially leading to higher attack success rates. This raises significant security and privacy concerns for existing AI systems. For example, privacy breaches related to user data or a company's proprietary AI model can be exploited through membership inference attacks [283] and model inversion attacks [284]. One potential approach for addressing privacy concerns in XAI is demonstrated by de Araújo [285], who employed Generative Adversarial Networks (GANs) to preserve the privacy of patient data while retaining important information about the optic nerve to diagnose Glaucoma using AI. Furthermore, Goethals et al. [286] proposed k-anonymous counterfactual explanations as a means to defend against privacy breaches arising from instance-based counterfactual algorithms utilizing nearest unlike neighbors [287].

Thus, the successful integration of XAI into 6G systems relies heavily on achieving a comprehensive trade-off among interpretability, performance, security, and privacy [288]. Achieving this trade-off requires close engagement with stakeholders for each 6G application and technology. This trade-off must be established through close engagement with stakeholders for each 6G application and technology. An essential step in this process is to identify the relative importance that stakeholders place on interpretability, performance, security, and privacy. Additionally, it is essential to develop a set of quantifiable metrics to measure the trade-off within each of these dimensions.

VII. LESSONS LEARNED AND FUTURE RESEARCH DIRECTIONS

This section briefly summarizes the lessons learned from the topics discussed in the previous sections. These include the background of AI and XAI, major research projects, and standardization efforts related to XAI and 6G, the impact of XAI on typical 6G technical aspects and use cases, and limitations and challenges when developing XAI for 6G. 6G. Additionally, we examine the corresponding future research directions for each of these topics.

A. XAI TECHNIQUE

1) LESSONS LEARNED

Before 2010, AI scientists focused primarily on improving the models' accuracy. As a result, the complexity of the models sharply increased (from simple rule-based models like decision trees to DL now). It wasn't until after 2010, when ML models were used in areas that impacted humans, such as diagnosing cancer or approving bank credit, that concerns arose regarding the transparency of AI decisions. Specifically, it became important to ensure that decisions were not being made based on factors like a person's ethnicity or race. These concerns raised questions about why the system decides in a particular way. In the imminent 6G age, where various model devices can talk to each other, more AI decisions will be made at a much faster pace. The transparency and trustworthiness of responsible AI have to be considered formally so that experts from academia and industry can improve the overall technology ecosystem for the future user experience.

XAI is a promising set of technologies that provide transparency in the decision-making process behind the AI black box. Although XAI is still in its early stages, researchers have already learned an important lesson: there is a trade-off between interpretability and performance. When stakeholders need more explainability, AI system designers may have to compromise the quality of the prediction or classification results. XAI also has a very significant role in validating the model. Sometimes, non-related factors to the output can bias models and affect their predictions.

Another significant lesson to be aware of is that ML models are heavily dependent on the quality of the training dataset [289]. Incorrect AI predictions or classification can lead to massive losses for 6G stakeholders in both economic and non-economical (e.g., health and life) aspects. This highlights the importance of AI systems' robustness. Most of the existing solutions [289] focus on improving the AI robustness, which is to carefully design the data pipeline, including data collection, pre-processing, augmentation, and dimensionality reduction. This helps with reducing the error rates.

2) KEY RESEARCH PROBLEMS

Explainability is the cornerstone for professionals adopting AI models and validating the logic of the models. Some of the important questions that arise in this area are:

- Is it possible to improve the explainability techniques by creating an extra layer to translate into layman's terms the logic behind ML models so we can provide models with both accuracy and suitable explanations?
- Would the next generation of ML models fill the demands of current legislations such as GDPR which highlights the importance of the right to explanation? So people are entitled to explanations of the outputs' algorithms that affect an individual?
- To what extent will ML models expand their use to other domains by providing better explanations to increase the professionals' trustfulness?

3) PRELIMINARY SOLUTIONS

One of the most prominent lessons learned during several decades of research, development, and commercialization of AI is that the performance of AI, especially ML systems, in certain domains, is not as important as the explainability/interpretability of the model. And that a significant strategy to involve professionals in implementing ML models is to increase their trustworthiness in the model by providing explanations of the decisions models make [28]. These explanations can be in the shape of text, graphs, or by providing an interpretable model [25]. Currently, there are many technologies used to generate explanations such as LIME and SHAP. However, they have a few limitations like that they do not well with all the models (Lime does not work well with XGBoost) or that SHAP is very slow in some methods such as k-NN. LIME is in general faster than SHAP, but LIME's explanations based on linear models do not guarantee the same levels of consistency as SHAP.

4) FUTURE DIRECTIONS

The wide and deep convergence of XAI to the existing AI systems is foreseen to be increasingly important. Some promising future research directions for this deployment include are: how to measure the level of explainability, how to satisfy the explainability demands from multiple stakeholders, how to ensure high performance while still being able to provide a high level of explainability, and how to work collaboratively in a multi-disciplinary team (i.e., typically, ICT researchers with legal experts). Generating clear and consistent explanations for accurate models easily remains a future challenge for AI with great potential to make AI more transparent and accessible to humans. Another interesting research direction is about making more XAI methods for multi-variant time-series data [290], which are widely seen in future mobile networks with lots of temporal data automatically generated from various sensors and end-devices.

B. STANDARDIZATION AND RESEARCH PROJECTS

1) LESSONS LEARNED

XAI is becoming an exciting research area under 6G. For the moment, IEEE is leading the standardization activities related to XAI. Especially, the IEEE Computer Society/Artificial Intelligence Standards Committee (C/AISC) and IEEE Intelligence Society - Standards Committee (CIS/SC) are leading these tasks [291], [292]. In addition, the National Institute of Standards and Technology (NIST) has published a report on a set of principles that can be used to judge the explainability of AI decisions [293]. This report defines the four principles of XAI: *Explanation:* Ability to provide reasons for the outcomes of the system, *Meaningful:* The provided explanation should be understandable and meaningful to the users, *Explanation Accuracy:* The provided explanation should accurately describe the process of generating the outcome, and *Knowledge Limits:* The system should understand the cases which are not designed or approved to operate or are unable to operate reliably.

However, none of these XAI standardization activities is focused on 6G or communication networks. Thus, it is yet to initiate the more focused XAI standardization activities for B5G and 6G domains. Moreover, the current 6G SDOs, such as ETSI, and 3GPP, have not focused entirely on the XAI domain. However, legal frameworks for XAI have already been developed at the global level, including in the EU and USA.

Several reputable research projects for 6G using XAI have already started. Mainly, the EU H2020 funding program, EU Christ-era funding program, EU MSCA program, and U.S. DARPA program have funded many projects in the XAI domain. Many of these projects are not directly related to 6G. However, most of these projects focus on technologies and applications associated with B5G and 6G networks.

2) KEY CHALLENGES

SDOs and funding organizations must address the following key challenges to support the integration of XAI into the 6G domain:

- How to establish collaborations between telecommunication SDOs, AI organizations, and funding agencies to define basic XAI requirements in 6G service deployment?
- Promote the development of open-source XAI projects and encourage funding organizations to invest in projects that enable XAI in the telecom sector.
- Provide training and educational programs to enhance the XAI expertise of both SDOs and funding agencies.
- How to foster partnerships between SDOs and funding organizations with leading XAI experts and companies to facilitate the integration of XAI into 6G services.

By tackling these challenges, SDOs and funding organizations can pave the way for the successful integration of XAI in the 6G domain, ensuring secure, accountable, and responsible AI-enabled telecom services.

3) PRELIMINARY SOLUTIONS

Standardization of Explainable AI (XAI) can be integrated with current standardization efforts for Zero-Touch Service and Network Management (ZSM). The focus of ZSM standardization is AI-powered service management in B5G networks. Leading telecom SDOs, such as ETSI, NGMN, 3GPP, and ITU-T, should consider incorporating XAI into their 6G standardization plans.

4) FUTURE DIRECTION

Given the substantial role that AI will play in 6G networks, it becomes imperative to assess the need for XAI in 6G applications. Initially, research projects can build new knowledge on utilizing XAI for B5G and 6G networks. EU funding programs such as Horizon Europe and Eureka programs can be ideal venues for funding research related to XAI and 6G integration. In addition, global level 6G programs such as Japan 6G/B5G Promotion Strategy and South Korea MSIT 6G research program can also be possible venues to obtain research funding for XAI and 6G integration. 6G standardization is essential to define the technological requirements of 6G networks and select suitable technologies to deploy 6G networks. XAI will be considered one of the critical technologies to utilize in 6G networks.

C. XAI FOR 6G TECHNICAL ASPECTS

1) LESSONS LEARNED

Conventional AI/ML algorithms and innovative DL architectures have been applied for different tasks in 6G networks when considering technical aspects. The objective includes accuracy improvement in intelligent radio and edge networks, reliability enhancement in network security and data privacy, and optimization in resource management. While the system performance and automatic decision of communication networks mostly depend on AI models, it does not usually provide descriptions and explanations about results, especially from the how-why-when perspective. XAI, owning to three principal features, including explainability, interpretability, and accountability, represents a promising technique to help not only end-users but also AI stakeholders understand how an AI model processes data and conducts outcomes automatically, which in turn allows end-users to be confident with its decision as well as engineers to comprehend their systems. Some ML models present good interpretability, but their performance in terms of accuracy is unacceptable. Therefore, the balance between interpretability and accuracy should be considered in XAI-based system design. For example, some XAI approaches have exploited the interpretability of AI models like a rule-based model and linear regression to generate explanations. However, their accuracy can not satisfy the baseline QoS in 6G. On the other hand, although DL showed high performance in dealing with many fundamental tasks such as detection, classification, and recognition, they offered little or no interpretability. Furthermore, depending on the input data type, storage infrastructure, computing platform, and communication infrastructure, XAI for explanation generation should be appropriately chosen to deal with a specific technical problem while ensuring a reasonable performance in terms of accuracy and complexity. Besides, the explanation should be simple for end-users with less domain knowledge and advanced for AI stakeholders with high expertise, which can be numerical results, text, graphs, images, and simulations. It can contain details on how a statistical AI model causes a prediction from a feature set, a decision path from a decision tree, a rule from a simple model, and a visual operation graph of information flow. Lastly, explanations brought by XAI may bring extra information leakage unintentionally to potential privacy violators. XAI can help strengthen AI accuracy and efficiency, but it can also tell others how existing AI models work, which in some cases should be confidential. Thus, extra attention should be put when applying XAI to privacy-based 6G solutions.

2) KEY RESEARCH PROBLEMS

To integrate XAI with 6G, a few challenges have to be addressed which are discussed below.

- Intelligent radio: With its interpretability and explainability, XAI has the potential to revolutionize intelligent radio in 6G, but it also increases the system's complexity. This creates a research problem on how to enhance explainability and interoperability, without increasing system complexity.
- Trust and security: Though XAI will provide administrators and stakeholders with significantly insightful, comprehensive security and trust information, the data may also be altered, which will influence the XAI model's decisions. This raises a research problem regarding the development of tamper-resistant storage and sharing of data.
- Privacy: Accountability will be improved by XAI, but the privacy of the shared data will be a problem since the data might be collected by a third party without the consent of the legitimate user.
- Resource management: XAI can help with resource management in 6G, the algorithms used for resource management cannot be explained at a high level because results vary from user to user. Context dependency is a research issue that needs to be addressed.
- Edge AI: XAI can improve the performance and explainability, but the issues, including the decreased performance of the AI algorithms and the absence of metrics to measure the performance of the XAI algorithms, must be addressed for 6G to be enabled by XAI with edge AI.
- Network automation and ZSM: XAI can improve the interpretability and justification of the decisions made by AI. However, performance deterioration of AI algorithms due to the integration of XAI is a concern that has to be addressed since the judgments made by ZSM may be mission-critical and may influence the network bandwidth and resource allocation.

3) PRELIMINARY SOLUTIONS

A critical component for reducing system complexity is a model governance environment. A good model of governance reduces the risk of a compliance audit and establishes the platform for transparent, ethical AI that eliminates bias in 6G networks. **black**Utilizing blockchain for data storage, later fed to XAI for decision-making in 6G networks, will enhance trust and security. The blockchain with its consensus, cryptography, and decentralization principles helps XAI to train on data that is tamper-resistant and trustworthy. By eliminating the need for data from local models in the creation of the global model, Federated Learning will improve the privacy of XAI in 6G networks. The use of incentives based on XAI will enhance resource sharing in 6G networks and help ZSM respond faster in mission-critical situations.

4) FUTURE DIRECTION

Despite certain benefits of interpretability and explainability, the utilization and development of XAI for different technical aspects in 6G networks are still limited. In this context, future work can focus on incorporating DL with several explanation techniques for XAI at multiple levels, from processing units to operation modules and systems. For example, the visual explanation technique can be applied to capture and then visualize the feature activation maps of a trained CNN, which helps explain the labels outputted by a DL-based automatic modulation classifier in intelligent radio. Besides visual explanation techniques (e.g., class activation mapping, peak response map, and class-enhanced attentive response), some other textual explanations (e.g., question-answering and semantic information retrieval) and numerical explanations (e.g., concept activation vectors and local interpretable model agnostic) approaches should be leveraged for a broad spectrum of 6G technical aspects. The combination of different explanation techniques (such as numerical and visual explanations) can be helpful for a sequence of interactive decisions produced by hierarchical AI models in complex systems. Many existing XAI works have concentrated on explainability for ML at different stages in developing AI systems to address several technical tasks in 6G. However, there remain some gaps in data explanation methods, which help select and explore a better-suited model later. Additionally, XAI methods should be designed to incorporate domain knowledge to explain useful inferences under clear and uncertain circumstances. One last promising future research direction is to identify to what certain level of explainability provided by XAI could be potentially harmful to preserving the privacy of XAI stakeholders.

D. XAI FOR 6G USE CASES

1) LESSONS LEARNED

Existing AI methodologies can provide prediction/classification for future 6G-based applications such as healthcare, Industry 5.0, CAV, smart grid, multi-sensory XR applications, and smart governance to help make decisions in real-time. Decision-taking in mission-critical applications such as healthcare, smart grid, and smart governance should be done very carefully as it may result in the loss of properties and lives and cause significant danger. However, the black-box nature of AI-based algorithms makes it very difficult for decision-makers to trust the results of these algorithms as they lack justification/explanation. The explanation should be technologically aware and thoroughly address the ethical, legal, and societal questions. To address these issues, XAI will be essential in future 6G-based applications (especially healthcare, autonomous driving, and smart governance) to trust, understand, and improve the accountability of the decisions made by AI-based algorithms. It will help instill confidence in end-users as they can understand the decision-making process of these algorithms. However, several key challenges and open issues need to be addressed to realize the full potential of XAI in the 6G-based applications which are discussed below.

The improved interpretability may result in reduced performance of AI algorithms in terms of real-time decision-making and prediction accuracy, which is unacceptable in mission-critical applications such as smart healthcare, autonomous vehicles, and smart grid. Hence, the trade-off between interpretability and performance is an open issue that needs addressing. Another important challenge is addressing the issue of the high dimensionality of the data generated from the applications based on IoT in real-time due to the high bandwidth and reduced latency of the 6G network infrastructure. Furthermore, the generation of labels for the data in real-time in the big data era makes it suitable for classification which is a tedious and demanding task. In the case of 6G-enabled applications that use heterogeneous networks, providing explainable and customized decisions is another open issue that needs to be addressed in future research. Another critical issue is the privacy preservation of sensitive data generated from applications such as healthcare, connected and autonomous vehicles, and smart grids. The malicious users or attackers can gain access to the private and sensitive data generated from these 6G applications through several means such as poisonous attacks.

2) KEY RESEARCH PROBLEMS

There are still some challenges that need to be addressed for the integration of XAI with 6G networks.

- In the case of intelligent health and wearables, and body area networks, healthcare stakeholders may benefit from explanations and assistance from XAI in interpreting AI models' decisions. However, the information used to feed XAI models may potentially come from unreliable sources, and the information from these sources can yield inaccurate results. Therefore, the decisions taken by stakeholders can have grave consequences. Identifying these unreliable sources is a research problem to be addressed.
- In the case of industry 5.0, collaborative robots, and digital twins, the stakeholders will benefit from the trustworthiness improvement, transparency enhancement, and result interaction provided by XAI. However, the process of decision-making requires data from multiple sensors, any fault in these sensors can lead to erroneous decisions, identifying the fault sensor in real time is an issue to be addressed.
- In the case of connected autonomous vehicles and UAVs, the drivers will benefit from the suggestions provided by XAI on issues related to collision alerts, driving alerts, and navigation assistance. However, most drivers lack the knowledge necessary to comprehend the justification and evaluate the decisions provided by XAI. Because of this issue, the aid provided by XAI is rendered ineffective in some circumstances.
- In the case of smart grid 2.0, the accountability provided by XAI will help identify the theft, and the reason for electric outrage, and also help the experts respond in an emergency. However, the complexity of the system will increase as it becomes necessary to get data from the whole chain of operations from multiple sources to produce decisions, which is an issue to be addressed.
- In the case of multi-sensory XR applications, holographic telepresence, and the metaverse, XAI can improve the quality of service and experience in these use cases. However, the issue related to system complexity, security, and privacy is still an issue to be addressed.
- In the case of smart governance, XAI can provide accountability and transparency for decisions made. However, due to politicization, there is a chance of possible conflict and the chance of having faulty outcomes, which is an issue that needs to be addressed. There is also a need for standards and guidelines for the integration of XAI with 6G.

3) PRELIMINARY SOLUTIONS

Many factors influence the decision-making of XAI models in 6G networks. The security and privacy of data are crucial factors that can be improved by combining blockchain and federated learning with XAI, which will also increase the trust in XAI decisions. In the use cases relating to 6G networks, a robust governance model can minimize bias, promote transparency, and decrease the chance of erroneous outcomes.

4) FUTURE DIRECTION

Some of the potential research directions that can address the aforementioned challenges and open issues are as follows. Researchers should focus on developing XAI algorithms that maintain the balance between explainability and the performance of the AI/ML algorithms by using technologies such as techno-economic analysis [294], [295]. Several soft computing techniques such as meta-heuristic algorithms, principal component analysis, and fuzzy systems can be considered to address the challenge of high dimensionality through dimensionality reduction [296]. Unsupervised learning algorithms such as clustering that do not require labels for prediction/classification can be used to address the issue of generation of labels in real-time for 6G-based applications [125]. Federated learning (FL), which is a recent development of ML, can be adopted in XAI-enabled 6G applications to provide customized decisions to heterogeneous networks [297]. Furthermore, FL can be integrated with XAI-enabled 6G applications to address the issue of privacy preservation [298], [299].

E. LIMITATIONS AND CHALLENGES OF XAI FOR 6G

1) LESSONS LEARNED

Recent studies on XAI methods in 6G exhibit three limitations. Firstly, there are not enough in-model XAI methods proposed so far. Most of the existing XAI can only explain black boxes after the 6G AI decision-making results are given. It prevents the achievement of a higher-level trade-off between the interpretability and model performance in 6G. Secondly, although many research studies have emphasized the importance of XAI measurements, there are no widely recognized quantifiable metrics for explainability in typical AI applications in 6G. Thirdly, there is a lack of multidisciplinary collaborations between experts in AI and the legal community. Finally, the increased transparency brought by XAI may lead to security and privacy concerns.

2) KEY RESEARCH PROBLEMS

XAI can be one of the engines to improve the development of 6G in the coming era. However, there are still a few challenges that need to be solved.

- Is it possible to develop better in-model XAI technologies for 6G to achieve a higher level of explainability and high decision-making performance?
- Will researchers be able to apply recognized metrics to evaluate explainability in 6G for end-users and stakeholders?
- How can XAI methods in 6G by engaging legal experts and experts from other disciplines meet the demands of current legislation and user satisfaction in all major stages (from design to evaluation)?
- What security and privacy threats can be posed by the increased transparency of AI decision-making?

3) PRELIMINARY SOLUTIONS

According to recent literature [16], most of the core systems that compose the wireless communications (signal detection, antenna detection, channel estimation, power allocation, etc.) which are essential in 6G have low, very low, or none explainability. Significant efforts are needed to enhance the explainability of 6G systems, particularly in critical domains like autonomous driving or remote surgery, to build trust among users. There are currently some frameworks proposed for integrating XAI in 6G and future wireless networks to help the understanding of the system by users and engineers in charge of the network infrastructure. These frameworks also consider malicious attacks from external threats [16].

DARPA addressed the lack of explainability in the U.S. in 2017. DARPA launched an initiative to promote XAI techniques to explain to humans the decisions taken by ML models [300]. This new challenge involved different areas such as designing explainable interfaces, understanding how the human mind understands concepts, and learning new explainable models. This was a four-year project and there were two different teams. One team is the XAI developers that worked on creating new effective techniques to provide useful explanations based on Human-computer interaction. The other team created an evaluation framework based on psychology to test the quality of the explanations. 6G can be used for very critical services such as remote surgery or autonomous driving [16]. The forthcoming EU AI Act [301] is anticipated to mark a significant milestone in the regulation of AI, aimed at addressing varying levels of risks associated with its usage. The introduction of the EU AI Act is likely to foster greater collaboration among all stakeholders involved in XAI 6G, including end-users and service providers, as they collectively confront challenges. For instance, the enforcement of this legislation would compel service providers to enhance the transparency of their AI solutions, thereby instilling greater trustworthiness among end-users.

4) FUTURE DIRECTION

In the upcoming years, we expect to see more applications of XAI in filling the gaps within the existing AI-driven 6G use cases and technical aspects. For example, as one type of quantum computing, the power of adiabatic quantum computation was validated in a 6G smart transportation pilot project for assigning optimal bus routes in the city of Lisbon, Portugal [302]. Such quantum computing-powered AI decision-making will be exponentially faster in the 6G ages due to more input data every second. It requires XAI technology to explain high-stakes decisions under strict performance pressure where data flows are extremely high in volume. Blockchain 3.0 [303] encompasses non-cryptocurrency blockchain applications such as electronic voting and supply chain management. In the 6G era, many heterogeneous blockchain systems need to be connected, which poses a great challenge to balancing network performance and system security and privacy demands. Moreover, the high heterogeneity of blockchain 3.0 also implies more diverse stakeholders from different organizations involved in 6G AI-assisted decision-making. Ensuring stakeholder satisfaction and compliance with local regulations for XAI methods would remain vital.

VIII. CONCLUSION

This paper provides a comprehensive review and analysis of the potential of using XAI methods to increase transparency and trustworthiness in a future AI-based 6G system. The paper begins with an overview of existing ideas for designing 6G networks, followed by an exhaustive survey of state-of-the-art AI and XAI methods. Several representative 6G technical aspects and use cases are then carefully analyzed, examining their existing AI-based solutions and the trend of applying XAI to enhance the trustworthiness of 6G network systems. Finally, the paper summarizes lessons learned about the limitations of existing work, reminding researchers and practitioners that XAI cannot solve all problems. The paper also highlights research challenges that show promise in overcoming or alleviating the potential limitations of XAI. The hope is that this survey will guide future 6G developments in a more sustainable direction.

REFERENCES

[1] K. David and H. Berndt, "6G vision and requirements: Is there any need for beyond 5G?" *IEEE Veh. Technol. Mag.*, vol. 13, no. 3, pp. 72–80, Sep. 2018.

[2] C. Benzaid and T. Taleb, "AI-driven zero touch network and service management in 5G and beyond: Challenges and research directions," *IEEE Netw.*, vol. 34, no. 2, pp. 186–194, Mar./Apr. 2020.

[3] L. Zhu, Z. Xiao, X.-G. Xia, and D. O. Wu, "Millimeter-wave communications with non-orthogonal multiple access for B5G/6G," *IEEE Access*, vol. 7, pp. 116123–116132, 2019.

[4] N. Kato et al., "Optimizing space-air-ground integrated networks by artificial intelligence," *IEEE Wireless Commun.*, vol. 26, no. 4, pp. 140–147, Aug. 2019.

[5] K. B. Letaief, W. Chen, Y. Shi, J. Zhang, and Y.-J. A. Zhang, "The roadmap to 6G: AI empowered wireless networks," *IEEE Commun. Mag.*, vol. 57, no. 8, pp. 84–90, Aug. 2019.

[6] S. Dang, O. Amin, B. Shihada, and M.-S. Alouini, "What should 6G be?" *Nat. Electron.*, vol. 3, no. 1, pp. 20–29, 2020.

[7] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *IEEE Netw.*, vol. 34, no. 3, pp. 134–142, May/Jun. 2020.

[8] Z. Zhang et al., "6G wireless networks: Vision, requirements, architecture, and key technologies," *IEEE Veh. Technol. Mag.*, vol. 14, no. 3, pp. 28–41, Sep. 2019.

[9] H. Sami, H. Otrok, J. Bentahar, and A. Mourad, "AI-based resource provisioning of IoE services in 6G: A deep reinforcement learning approach," *IEEE Trans. Netw. Service Manag.*, vol. 18, no. 3, pp. 3527–3540, Sep. 2021.

[10] S. Saab Jr., K. Saab, S. Phoha, M. Zhu, and A. Ray, "A multivariate adaptive gradient algorithm with reduced tuning efforts," *Neural Netw.*, vol. 152, pp. 499–509, Aug. 2022.

[11] M. Liyanage et al., "A survey on zero touch network and service (ZSM) management for 5G and beyond networks," *J. Netw. Comput. Appl.*, vol. 203, Jul. 2022, Art. no. 103362.

[12] S. K. Singh, R. Singh, and B. Kumbhani, "The evolution of radio access network towards open-RAN: Challenges and opportunities," in *Proc. IEEE Wireless Commun. Netw. Conf. Workshops (WCNCW)*, 2020, pp. 1–6.

[13] H. Yang, A. Alphones, Z. Xiong, D. Niyato, J. Zhao, and K. Wu, "Artificial-intelligence-enabled intelligent 6G networks," *IEEE Netw.*, vol. 34, no. 6, pp. 272–280, Nov./Dec. 2020.

[14] D. Slack, S. Hilgard, E. Jia, S. Singh, and H. Lakkaraju, "Fooling LIME and SHAP: Adversarial attacks on post hoc explanation methods," in *Proc. AAAI/ACM Conf. AI, Ethics, Soc.*, 2020, pp. 180–186.

[15] D. Kaur, S. Uslu, A. Durresi, S. Badve, and M. Dundar, "Trustworthy explainability acceptance: A new metric to measure the trustworthiness of interpretable AI medical diagnostic systems," in *Proc. Conf. Complex, Intell., Softw. Intensive Syst.*, 2021, pp. 35–46.

[16] W. Guo, "Explainable artificial intelligence for 6G: Improving trust between human and machine," *IEEE Commun. Mag.*, vol. 58, no. 6, pp. 39–45, Jun. 2020.

[17] T. D. Grant and D. J. Wischik, "Show us the data: Privacy, explainability, and why the law can't have both," *George Washington Law Rev.*, vol. 88, p. 1350, Dec. 2020.

[18] S. Qiu, Q. Liu, S. Zhou, and C. Wu, "Review of artificial intelligence adversarial attack and defense technologies," *Appl. Sci.*, vol. 9, no. 5, p. 909, 2019.

[19] D. Gunning and D. Aha, "DARPA's explainable artificial intelligence (XAI) program," *AI Mag.*, vol. 40, no. 2, pp. 44–58, 2019.

[20] A. Das and P. Rad, "Opportunities and challenges in explainable artificial intelligence (XAI): A survey," 2020, *arXiv:2006.11371*.

[21] M. E. Morocho-Cayamcela, H. Lee, and W. Lim, "Machine learning for 5G/B5G mobile and wireless communications: Potential, limitations, and future directions," *IEEE Access*, vol. 7, pp. 137184–137206, 2019.

[22] P. Porambage, G. Gür, D. P. M. Osorio, M. Liyanage, A. Gurtov, and M. Ylianttila, "The roadmap to 6G security and privacy," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 1094–1122, 2021.

[23] C. Li, W. Guo, S. C. Sun, S. Al-Rubaye, and A. Tsourdos, "Trustworthy deep learning in 6G-enabled mass autonomy: From concept to quality-of-trust key performance indicators," *IEEE Veh. Technol. Mag.*, vol. 15, no. 4, pp. 112–121, Dec. 2020.

[24] A. Adadi and M. Berrada, "Peeking inside the black-box: A survey on explainable artificial intelligence (XAI)," *IEEE Access*, vol. 6, pp. 52138–52160, 2018.

[25] A. B. Arrieta et al., "Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Inf. Fusion*, vol. 58, pp. 82–115, Jun. 2020.

[26] L. Arras, A. Osman, K.-R. Müller, and W. Samek, "Evaluating recurrent neural network explanations," 2019, *arXiv:1904.11829*.

[27] M. T. Ribeiro, S. Singh, and C. Guestrin, "Model-agnostic interpretability of machine learning," 2016, *arXiv:1606.05386*.

[28] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should I trust you?' Explaining the predictions of any classifier," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Disc. Data Min.*, 2016, pp. 1135–1144.

[29] W. Samek and K.-R. Müller, "Towards explainable artificial intelligence," in *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*. Cham, Switzerland: Springer, 2019, pp. 5–22.

[30] "ELI5: A library for debugging/inspecting machine learning classifiers and explaining their predictions." Team ELI5. 2021. [Online]. Available: https://github.com/TeamHG-Memex/eli5

[31] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Stat.*, vol. 29, no. 5, pp. 1189–1232, 2001.

[32] N. Papernot and P. McDaniel, "Deep k-nearest neighbors: Towards confident, interpretable and robust deep learning," 2018, *arXiv:1803.04765*.

[33] M. Hind et al., "TED: Teaching AI to explain its decisions," in *Proc. AAAI/ACM Conf. AI, Ethics, Soc.*, 2019, pp. 123–129.

[34] Q. Zhang, Y. N. Wu, and S.-C. Zhu, "Interpretable convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 8827–8836.

[35] L. S. Shapley, "A value for n-person games," in *Contributions to Theory of Games*. Santa Monica, CA, USA: RAND Corp., 1953, pp. 307–317.

[36] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 4768–4777.

[37] S. M. Lundberg, G. G. Erion, and S.-I. Lee, "Consistent individualized feature attribution for tree ensembles," 2018, *arXiv:1802.03888*.

[38] N. Jethani, M. Sudarshan, I. C. Covert, S.-I. Lee, and R. Ranganath, "FastSHAP: Real-time Shapley value estimation," in *Proc. Int. Conf. Learn. Represent.*, 2021, pp. 1–20.

[39] G. Montavon, A. Binder, S. Lapuschkin, W. Samek, and K.-R. Müller, "Layer-wise relevance propagation: An overview," in *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*, Cham, Switzerland: Springer, pp. 193–209, 2019.

[40] 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 *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 618–626.

[41] S. Sharma, J. Henderson, and J. Ghosh, "CERTIFAI: Counterfactual explanations for robustness, transparency, interpretability, and fairness of artificial intelligence models," 2019, *arXiv:1905.07857*.

[42] S. Wachter, B. Mittelstadt, and C. Russell, "Counterfactual explanations without opening the black box: Automated decisions and the GDPR," *Harvard J. Law Technol.*, vol. 31, no. 2, p. 841, 2018.

[43] I. Ben-Gal, "Bayesian networks," in *Encyclopedia of Statistics in Quality and Reliability*, vol. 1. Chichester, U.K.: Wiley, 2008.

[44] A. Fisher, C. Rudin, and F. Dominici, "All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously," *J. Mach. Learn. Res.*, vol. 20, no. 177, pp. 1–81, 2019.

[45] O. Ayoub, F. Musumeci, F. Ezzeddine, C. Passera, and M. Tornatore, "On using explainable artificial intelligence for failure identification in microwave networks," in *Proc. 25th Conf. Innov. Clouds, Internet Netw. (ICIN)*, 2022, pp. 48–55.

[46] M. Ribera and A. Lapedriza, "Can we do better explanations? A proposal of user-centered explainable AI," in *Proc. IUI Workshops*, 2019, p. 38.

[47] M. R. Wick and W. B. Thompson, "Reconstructive explanation: Explanation as complex problem solving," in *Proc. IJCAI*, 1989, pp. 135–140.

[48] M. A. Qureshi and D. Greene, "EVE: Explainable vector based embedding technique using Wikipedia," *J. Intell. Inf. Syst.*, vol. 53, no. 1, pp. 137–165, 2019.

[49] M. A. Qureshi and D. Greene, "Lit@ EVE: Explainable recommendation based on Wikipedia concept vectors," in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discovery Databases*, 2017, pp. 409–413.

[50] "General data protection regulation (GDPR) compliance." Accessed: Oct. 1, 2023. [Online]. Available: https://gdpr.eu/

[51] "summary of the HIPAA security rule." Accessed: Oct. 1, 2023. [Online]. Available: https://www.hhs.gov/hipaa/for-professionals/security/laws-regulations/index.html

[52] "Gramm-leach-Bliley act." Accessed: Oct. 1, 2023. [Online]. Available: https://www.ftc.gov/tips-advice/business-center/privacy-and-security/gramm-leach-bliley-act

[53] "S.2521—Federal information security modernization act of 2014." Accessed: Oct. 1, 2023. [Online]. Available: https://www.congress.gov/bill/113th-congress/senate-bill/2521

[54] "Protecting controlled unclassified information in nonfederal information systems and organizations." Accessed: Oct. 1, 2023. [Online]. Available: http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-171.pdf

[55] "China introduces first comprehensive legislation on personal information protection." Accessed: Oct. 1, 2023. [Online]. Available: https://www.lw.com/thoughtLeadership/china-introduces-first-comprehensive-legislation-on-personal-information-protection

[56] "What changes with the LGPD." Accessed: Oct. 1, 2023. [Online]. Available: https://www.serpro.gov.br/lgpd/menu/a-lgpd/o-que-muda-com-a-lgpd

[57] "Notifiable data breaches." Accessed: Oct. 1, 2023. [Online]. Available: https://www.oaic.gov.au/privacy/notifiable-data-breaches/

[58] "Personal information protection commission." Accessed: Oct. 15, 2023. [Online]. Available: https://www.ppc.go.jp/en/

[59] "Horizon 2020." Accessed: Oct. 1, 2023. [Online]. Available: https://ec.europa.eu/programmes/horizon2020/en/home

[60] "AI4EU." Accessed: Oct. 1, 2023. [Online]. Available: https://www.ai4eu.eu/

[61] "FeatureCloud." Accessed: Oct. 1, 2023. [Online]. Available: https://featurecloud.eu/

[62] "Explainable manufacturing artificial intelligence (XMANAI)." Accessed: Oct. 1, 2023. [Online]. Available: https://ai4manufacturing.eu/

[63] "Explainable AI pipelines for big copernicus data (DEEPCUBE)." Accessed: Oct. 1, 2023. [Online]. Available: https://deepcube-h2020.eu/

[64] "Security and privacy accountable technology innovations, algorithms, and machine learning (SPATIAL)." Accessed: Oct. 1, 2023. [Online]. Available: https://cordis.europa.eu/project/id/101021808

[65] "Safe and trusted human centric artificial intelligence in future manufacturing lines (STAR)." Accessed: Oct. 1, 2023. [Online]. Available: https://star-ai.eu/

[66] "Confidential computing and privacy-preserving technologies for 6G (confidential 6G)." Accessed: Oct. 1, 2023. [Online]. Available: https://confidential6g.eu/

[67] "Data aware wireless networks for Internet of Everything (DAWN4IoE)." Accessed: Oct. 1, 2023. [Online]. Available: https://cordis.europa.eu/project/id/778305

[68] "European level 6G flagship project (Hexa-X-II)." Accessed: Oct. 1, 2023. [Online]. Available: https://hexa-x-ii.eu/

[69] "Copernicus:europe's eyes on earth." Accessed: Oct. 1, 2023. [Online]. Available: https://www.copernicus.eu/en

[70] "Marie Skłodowska-curie actions." Accessed: Oct. 1, 2023. [Online]. Available: https://ec.europa.eu/programmes/horizon2020/en/h2020-section/marie-sklodowska-curie-actions

[71] "Interactive natural language technology for explainable artificial intelligence (NL4XAI)." Accessed: Oct. 1, 2023. [Online]. Available: https://nl4xai.eu/

[72] "Building greener and more sustainable societies by filling the knowledge gap in social science and engineering to enable responsible artificial intelligence co-creation (GECKO)." Accessed: Oct. 1, 2023. [Online]. Available: https://gecko-project.eu/

[73] "Explainable artificial intelligence (XAI)." Accessed: Oct. 1, 2023. [Online]. Available: https://www.darpa.mil/program/explainable-artificial-intelligence

[74] "Explainable artificial intelligence(XAI) Center." Accessed: Oct. 1, 2023. [Online]. Available: http://xai.kaist.ac.kr/

[75] "Secure design and deployment of trusthworthy continuum computing 6G services (RIGOUROUS)." Accessed: Oct. 1, 2023. [Online]. Available: https://rigourous.eu/

[76] K. Cao et al., "Improving physical layer security of uplink NOMA via energy harvesting jammers," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 786–799, 2021.

[77] "O-RAN AI/ML workflow description and requirements 1.03," O-RAN Working Group 2, O-RAN.WG2.AIML-v01.03 Technical Specification, O-RAN Alliance e.V., Alfter, Germany, Jul. 2021.

[78] M. Polese, L. Bonati, S. D'Oro, S. Basagni, and T. Melodia, "ColO-RAN: Developing machine learning-based xApps for open RAN closed-loop control on programmable experimental platforms," *IEEE Trans. Mobile Comput.*, vol. 22, no. 10, pp. 5787–5800, Oct. 2023.

[79] Z. Chen, J. Tang, X. Y. Zhang, D. K. C. So, S. Jin, and K.-K. Wong, "Hybrid evolutionary-based sparse channel estimation for IRS-assisted mmWave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 21, no. 3, pp. 1586–1601, Mar. 2022.

[80] Q.-V. Pham, N. T. Nguyen, T. Huynh-The, L. B. Le, K. Lee, and W.-J. Hwang, "Intelligent radio signal processing: A survey," *IEEE Access*, vol. 9, pp. 83818–83850, 2021.

[81] T. Huynh-The, C.-H. Hua, Q.-V. Pham, and D.-S. Kim, "MCNet: An efficient CNN architecture for robust automatic modulation classification," *IEEE Commun. Lett.*, vol. 24, no. 4, pp. 811–815, Apr. 2020.

[82] P. Hũu, M. A. Arfaoui, S. Sharafeddine, C. M. Assi, and A. Ghrayeb, "A low-complexity framework for joint user pairing and power control for cooperative NOMA in 5G and beyond cellular networks," *IEEE Trans. Commun.*, vol. 68, no. 11, pp. 6737–6749, Nov. 2020.

[83] T. Huynh-The et al., "Automatic modulation classification: A deep architecture survey," *IEEE Access*, vol. 9, pp. 142950–142971, 2021.

[84] M. Polese, L. Bonati, S. D'Oro, S. Basagni, and T. Melodia, "Understanding O-RAN: Architecture, interfaces, algorithms, security, and research challenges," 2022, *arXiv:2202.01032*.

[85] X.-Q. Pham, T.-D. Nguyen, T. Huynh-The, E.-N. Huh, and D.-S. Kim, "Distributed cloud computing: Architecture, enabling technologies, and open challenges," *IEEE Consum. Electron. Mag.*, vol. 12, no. 3, pp. 98–106, May 2023.

[86] G. B. Tunze, T. Huynh-The, J.-M. Lee, and D.-S. Kim, "Sparsely connected CNN for efficient automatic modulation recognition," *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 15557–15568, Dec. 2020.

[87] L. J. Wong and S. McPherson, "Explainable neural network-based modulation classification via concept bottleneck models," in *Proc. IEEE 11th Annu. Comput. Commun. Workshop Conf. (CCWC)*, 2021, pp. 0191–0196.

[88] H.-S. Lee, "Channel metamodeling for explainable data-driven channel model," *IEEE Wireless Commun. Lett.*, vol. 10, no. 12, pp. 2678–2682, Dec. 2021.

[89] H. Guo and V. K. N. Lau, "Robust deep learning for uplink channel estimation in cellular network under inter-cell interference," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 6, pp. 1873–1887, Jun. 2023.

[90] S. D'Oro, M. Polese, L. Bonati, H. Cheng, and T. Melodia, "dApps: Distributed applications for real-time inference and control in O-RAN," *IEEE Commun. Mag.*, vol. 60, no. 11, pp. 52–58, Nov. 2022.

[91] L. Mucchi et al., "Physical-layer security in 6G networks," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 1901–1914, 2021.

[92] M. Liyanage, A. Braeken, A. D. Jurcut, M. Ylianttila, and A. Gurtov, "Secure communication channel architecture for software defined mobile networks," *Comput. Netw.*, vol. 114, pp. 32–50, Feb. 2017.

[93] P. Porambage, G. Gür, D. P. Moya Osorio, M. Livanage, and M. Ylianttila, "6G security challenges and potential solutions," in *Proc. Joint Eur. Conf. Netw. Commun. 6G Summit (EuCNC/6G Summit)*, 2021, pp. 622–627.

[94] F. Tang, Y. Kawamoto, N. Kato, and J. Liu, "Future intelligent and secure vehicular network toward 6G: Machine-learning approaches," *Proc. IEEE*, vol. 108, no. 2, pp. 292–307, Feb. 2020.

[95] Y. Siriwardhana, P. Porambage, M. Liyanage, and M. Ylianttila, "AI and 6G security: Opportunities and challenges," in *Proc. Joint Eur. Conf. Netw. Commun. 6G Summit (EuCNC/6G Summit)*, 2021, pp. 616–621.

[96] L. Viganò and D. Magazzeni, "Explainable security," in *Proc. IEEE Eur. Symp. Security Privacy Workshops (EuroS PW)*, 2020, pp. 293–300.

[97] M. Zolanvari, Z. Yang, K. Khan, R. Jain, and N. Meskin, "TRUST XAI: Model-agnostic explanations for AI with a case study on IIoT security," *IEEE Internet Things J.*, vol. 10, no. 4, pp. 2967–2978, Feb. 2023.

[98] Y. Chai, Y. Zhou, W. Li, and Y. Jiang, "An explainable multi-modal hierarchical attention model for developing phishing threat intelligence," *IEEE Trans. Dependable Secure Comput.*, vol. 19, no. 2, pp. 790–803, Mar./Apr. 2022.

[99] B. Gulmezoglu, "XAI-based microarchitectural side-channel analysis for Website fingerprinting attacks and defenses," *IEEE Trans. Dependable Secure Comput.*, vol. 19, no. 6, pp. 4039–4051, Nov./Dec. 2022.

[100] Y. Sun, J. Liu, J. Wang, Y. Cao, and N. Kato, "When machine learning meets privacy in 6G: A survey," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 4, pp. 2694–2724, 4th Quart., 2020.

[101] K. Cao et al., "On the ergodic secrecy capacity of intelligent reflecting surface aided wireless powered communication systems," *IEEE Wireless Commun. Lett.*, vol. 11, no. 11, pp. 2275–2279, Nov. 2022.

[102] H. Fu, Z. Zheng, S. Zhu, and P. Mohapatra, "Keeping context in mind: Automating mobile App access control with user interface inspection," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Paris, France, May 2019, pp. 2089–2097.

[103] X. Wu, L. Qi, J. Gao, G. Ji, and X. Xu, "An ensemble of random decision trees with local differential privacy in edge computing," *Neurocomputing*, vol. 485, pp. 181–195, May 2022.

[104] A. Dhurandhar, K. Shanmugam, R. Luss, and P. Olsen, "Improving simple models with confidence profiles," in *Proc. 32nd Int. Conf. Neural Inf. Process. Syst.*, Red Hook, NY, USA, 2018, pp. 10317–10327.

[105] B. Liu, M. Ding, S. Shaham, W. Rahayu, F. Farokhi, and Z. Lin, "When machine learning meets privacy: A survey and outlook," *ACM Comput. Surveys*, vol. 54, no. 2, p. 31, Mar. 2021.

[106] C. Xu, J. Ren, D. Zhang, Y. Zhang, Z. Qin, and K. Ren, "GANobfuscator: Mitigating information leakage under GAN via differential privacy," *IEEE Trans. Inf. Forensics Security*, vol. 14, pp. 2358–2371, 2019.

[107] J. Yoon, L. N. Drumright, and M. van der Schaar, "Anonymization through data synthesis using generative adversarial networks (ADS-GAN)," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 8, pp. 2378–2388, Aug. 2020.

[108] V. Nagisetty, L. Graves, J. Scott, and V. Ganesh, "xAI-GAN: Enhancing generative adversarial networks via explainable AI systems," 2020, *arXiv:2002.10438*.

[109] I. Ahmad et al., "Towards gadget-free Internet services: A roadmap of the naked world," *Telemat. Inform.*, vol. 35, no. 1, pp. 82–92, 2018.

[110] F. Hussain, S. A. Hassan, R. Hussain, and E. Hossain, "Machine learning for resource management in cellular and IoT networks: Potentials, current solutions, and open challenges," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 2, pp. 1251–1275, 2nd Quart., 2020.

[111] S. Yu, X. Chen, Z. Zhou, X. Gong, and D. Wu, "When deep reinforcement learning meets federated learning: Intelligent multi-timescale resource management for multiaccess edge computing in 5G ultradense network," *IEEE Internet Things J.*, vol. 8, no. 4, pp. 2238–2251, Feb. 2021.

[112] H. Zhang, H. Zhang, K. Long, and G. K. Karagiannidis, "Deep learning based radio resource management in NOMA networks: User association, subchannel and power allocation," *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 4, pp. 2406–2415, Oct.–Dec. 2020.

[113] M. Lin and Y. Zhao, "Artificial intelligence-empowered resource management for future wireless communications: A survey," *China Commun.*, vol. 17, no. 3, pp. 58–77, Mar. 2020.

[114] W. Guan, H. Zhang, and V. C. M. Leung, "Customized slicing for 6G: Enforcing artificial intelligence on resource management," *IEEE Netw.*, vol. 35, no. 5, pp. 264–271, Sep./Oct. 2021.

[115] S. Wang, H. Sheng, Y. Zhang, D. Yang, J. Shen, and R. Chen, "Blockchain-empowered distributed multicamera multitarget tracking in edge computing," *IEEE Trans. Ind. Informat.*, vol. 20, no. 1, pp. 369–379, Jan. 2024.

[116] A. Nascita, A. Montieri, G. Aceto, D. Ciuonzo, V. Persico, and A. Pescapé, "XAI meets mobile traffic classification: Understanding and improving multimodal deep learning architectures," *IEEE Trans. Netw. Service Manag.*, vol. 18, no. 4, pp. 4225–4246, Dec. 2021.

[117] C. Callegari, P. Ducange, M. Fazzolari, and M. Vecchio, "Explainable Internet traffic classification," *Appl. Sci.*, vol. 11, no. 10, p. 4697, 2021.

[118] Z. Ali, L. Jiao, T. Baker, G. Abbas, Z. H. Abbas, and S. Khaf, "A deep learning approach for energy efficient computational offloading in mobile edge computing," *IEEE Access*, vol. 7, pp. 149623–149633, 2019.

[119] J. Baek and G. Kaddoum, "Heterogeneous task offloading and resource allocations via deep recurrent reinforcement learning in partial observable multifog networks," *IEEE Internet Things J.*, vol. 8, no. 2, pp. 1041–1056, Jan. 2021.

[120] U. Schlegel, H. Arnout, M. El-Assady, D. Oelke, and D. A. Keim, "Towards a Rigorous evaluation of XAI methods on time series," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Seoul, South Korea, Oct. 2019, pp. 4197–4201.

[121] X. Wang, Y. Han, C. Wang, Q. Zhao, X. Chen, and M. Chen, "In-edge AI: Intelligentizing mobile edge computing, caching and communication by federated learning," *IEEE Netw.*, vol. 33, no. 5, pp. 156–165, Sep./Oct. 2019.

[122] Y. Xiao, G. Shi, Y. Li, W. Saad, and H. V. Poor, "Toward self-learning edge intelligence in 6G," *IEEE Commun. Mag.*, vol. 58, no. 12, pp. 34–40, Dec. 2020.

[123] Z. Zhao, Z. Ding, T. Q. Quek, and M. Peng, "Edge artificial intelligence in 6G systems: Theory, key techniques, and applications," *China Commun.*, vol. 17, no. 8, pp. 3–4, Aug. 2020.

[124] I. Tomkos, D. Klonidis, E. Pikasis, and S. Theodoridis, "Toward the 6G network era: Opportunities and challenges," *IT Prof.*, vol. 22, no. 1, pp. 34–38, Jan./Feb. 2020.

[125] J. Kaur, M. A. Khan, M. Iftikhar, M. Imran, and Q. E. U. Haq, "Machine learning techniques for 5G and beyond," *IEEE Access*, vol. 9, pp. 23472–23488, 2021.

[126] Y. Zhao, J. Zhao, W. Zhai, S. Sun, D. Niyato, and K.-Y. Lam, "A survey of 6G wireless communications: Emerging technologies," in *Proc. Future Inf. Commun. Conf.*, 2021, pp. 150–170.

[127] P. Porambage et al., "Sec-EdgeAI: AI for edge security vs security for edge AI," in *Proc. 1st 6G Wireless Summit*, Levi, Finland, 2019, pp. 1–2.

[128] Y. Li, Y. Yu, W. Susilo, Z. Hong, and M. Guizani, "Security and privacy for edge intelligence in 5G and beyond networks: Challenges and solutions," *IEEE Wireless Commun.*, vol. 28, no. 2, pp. 63–69, Apr. 2021.

[129] R. Shafin, L. Liu, V. Chandrasekhar, H. Chen, J. Reed, and J. C. Zhang, "Artificial intelligence-enabled cellular networks: A critical path to beyond-5G and 6G," *IEEE Wireless Commun.*, vol. 27, no. 2, pp. 212–217, Apr. 2020.
[130] C. Chaccour and W. Saad, "Edge intelligence in 6G systems," in *6G Mobile Wireless Networks*. Cham, Switzerland: Springer, 2021, pp. 233–249.
[131] M. S. Hossain, G. Muhammad, and N. Guizani, "Explainable AI and mass surveillance system-based healthcare framework to combat COVID-I9 like pandemics," *IEEE Netw.*, vol. 34, no. 4, pp. 126–132, Jul./Aug. 2020.
[132] M. Liyanage et al., "Enhancing security of software defined mobile networks," *IEEE Access*, vol. 5, pp. 9422–9438, 2017.
[133] M. Liyanage, I. Ahmad, M. Ylianttila, A. Gurtov, A. B. Abro, and E. M. De Oca, "Leveraging LTE security with SDN and NFV," in *Proc. IEEE 10th Int. Conf. Ind. Inf. Syst. (ICIIS)*, 2015, pp. 220–225.
[134] C. Benzaid and T. Taleb, "ZSM security: Threat surface and best practices," *IEEE Netw.*, vol. 34, no. 3, pp. 124–133, May/Jun. 2020.
[135] J. Ortiz et al., "INSPIRE-5Gplus: Intelligent security and pervasive trust for 5G and beyond networks," in *Proc. 15th Int. Conf. Availability, Rel. Security*, 2020, pp. 1–10.
[136] "AIOps (artificial intelligence for IT operations)." Accessed: Sep. 14, 2023. [Online]. Available: https://www.gartner.com/en/information-technology/glossary/aiops-artificial-intelligence-operations
[137] B. Dutta, A. Krichel, and M.-P. Odini, "The challenge of zero touch and explainable AI," *J. ICT Stand.*, vol. 9, no. 2, pp. 147–158, 2021.
[138] M. S. Kaiser et al., "6G access network for intelligent Internet of Healthcare Things: Opportunity, challenges, and research directions," in *Proc. Int. Conf. Trends Comput. Cogn. Eng.*, 2021, pp. 317–328.
[139] D. Bega, M. Gramaglia, A. Garcia-Saavedra, M. Fiore, A. Banchs, and X. Costa-Perez, "Network slicing meets artificial intelligence: An AI-based framework for slice management," *IEEE Commun. Mag.*, vol. 58, no. 6, pp. 32–38, Jun. 2020.
[140] P. Barnard, I. Macaluso, N. Marchetti, and L. P. da Silva, "Resource reservation in sliced networks: An explainable artificial intelligence (XAI) approach," in *Proc. ICC*, 2022, pp. 1530–1535.
[141] M. Giordani, M. Polese, M. Mezzavilla, S. Rangan, and M. Zorzi, "Toward 6G networks: Use cases and technologies," *IEEE Commun. Mag.*, vol. 58, no. 3, pp. 55–61, Mar. 2020.
[142] C. D. Alwis et al., "Survey on 6G frontiers: Trends, applications, requirements, technologies and future research," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 836–886, 2021.
[143] I. Bisio, C. Garibotto, F. Lavagetto, and A. Sciarrone, "When eHealth meets IoT: A smart wireless system for post-stroke home rehabilitation," *IEEE Wireless Commun.*, vol. 26, no. 6, pp. 24–29, Dec. 2019.
[144] A. R. Javed et al., "Toward explainable AI-empowered cognitive health assessment," *Front. Public Health*, vol. 11, Mar. 2023, Art. no. 1024195.
[145] S. Hossain, A. Chakrabarty, T. R. Gadekallu, M. Alazab, and M. J. Piran, "Vision transformers, ensemble model, and transfer learning leveraging explainable ai for brain tumor detection and classification," *IEEE J. Biomed. Health Inform.*, vol. 28, no. 3, pp. 1261–1272, Mar. 2024.
[146] A. Garg and V. Mago, "Role of machine learning in medical research: A survey," *Comput. Sci. Rev.*, vol. 40, May 2021, Art. no. 100370.
[147] N. Rieke et al., "The future of digital health with federated learning," *NPJ Digit. Med.*, vol. 3, no. 1, pp. 1–7, 2020.
[148] A. Qayyum, J. Qadir, M. Bilal, and A. Al-Fuqaha, "Secure and robust machine learning for healthcare: A survey," *IEEE Rev. Biomed. Eng.*, vol. 14, pp. 156–180, 2021.
[149] T. Huynh-The, C.-H. Hua, N. A. Tu, and D.-S. Kim, "Physical activity recognition with statistical-deep fusion model using multiple sensory data for smart health," *IEEE Internet Things J.*, vol. 8, no. 3, pp. 1533–1543, Feb. 2021.
[150] C.-H. Hua et al., "Convolutional network with twofold feature augmentation for diabetic retinopathy recognition from multi-modal images," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 7, pp. 2686–2697, Jul. 2021.
[151] H. Cao et al., "A two-stage convolutional neural networks for lung nodule detection," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 7, pp. 2006–2015, Jul. 2020.
[152] T. Nakamura, Y. D. Alqurashi, M. J. Morrell, and D. P. Mandic, "Hearables: Automatic overnight sleep monitoring with standardized in-ear EEG sensor," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 1, pp. 203–212, Jan. 2020.
[153] C.-H. Hua et al., "Bimodal learning via trilogy of skip-connection deep networks for diabetic retinopathy risk progression identification," *Int. J. Med. Inform.*, vol. 132, Dec. 2019, Art. no. 103926.
[154] X. Li, M. Jia, M. T. Islam, L. Yu, and L. Xing, "Self-supervised feature learning via exploiting multi-modal data for retinal disease diagnosis," *IEEE Trans. Med. Imag.*, vol. 39, no. 12, pp. 4023–4033, Dec. 2020.
[155] X. Liu, L. A. Finelli, G. L. Hersch, and I. Khalil, "Attention-based LSTM network for COVID-19 clinical trial parsing," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Atlanta, GA, USA, Dec. 2020, pp. 3761–3766.
[156] J. Hou et al., "Deep learning and data augmentation based data imputation for structural health monitoring system in multi-sensor damaged state," *Measurement*, vol. 196, Jun. 2022, Art. no. 111206.
[157] U. Pawar, D. O'Shea, S. Rea, and R. O'Reilly, "Explainable AI in healthcare," in *Proc. Int. Conf. Cyber Situat. Awareness, Data Anal. Assess. (CyberSA)*, Dublin, Ireland, Jun. 2020, pp. 1–2.
[158] P. A. Moreno-Sanchez, "An automated feature selection and classification pipeline to improve explainability of clinical prediction models," in *Proc. IEEE 9th Int. Conf. Healthcare Inform. (ICHI)*, Victoria, BC, Canada, Aug. 2021, pp. 527–534.
[159] T. A. Schoonderwoerd, W. Jorritsma, M. A. Neerincx, and K. van den Bosch, "Human-centered XAI: Developing design patterns for explanations of clinical decision support systems," *Int. J. Human-Comput. Stud.*, vol. 154, Jun. 2021, Art. no. 102684.
[160] N. Prentzas, M. Pitsiali, E. Kyriacou, A. Nicolaides, A. Kakas, and C. S. Pattichis, "Model agnostic explainability techniques in ultrasound image analysis," in *Proc. IEEE 21st Int. Conf. Bioinf. Bioeng. BIBE)*, Kragujevac, Serbia, Oct. 2021, pp. 1–6.
[161] Y. Yan, J. Zhu, M. Duda, E. Solarz, C. Sripada, and D. Koutra, "GroupiNN: Grouping-based interpretable neural network for classification of limited, noisy brain data," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Disc. Data Min.*, 2019, pp. 772–782.
[162] M. A. Gulum, C. M. Trombley, and M. Kantardzic, "Improved deep learning explanations for prostate lesion classification through grad-CAM and saliency map fusion," in *Proc. IEEE 34th Int. Symp. Comput.-Based Med. Syst. (CBMS)*, Aveiro, Portugal, Jun. 2021, pp. 498–502.
[163] A. Lucieri, M. N. Bajwa, S. A. Braun, M. I. Malik, A. Dengel, and S. Ahmed, "ExAID: A multimodal explanation framework for computer-aided diagnosis of skin lesions," *Comput. Methods Programs Biomed.*, vol. 215, Mar. 2022, Art. no. 106620.
[164] E. Tjoa and C. Guan, "A survey on explainable artificial intelligence (XAI): Toward medical XAI," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 11, pp. 4793–4813, Nov. 2021.
[165] B. C. Kwon et al., "RetainVis: Visual analytics with interpretable and interactive recurrent neural networks on electronic medical records," *IEEE Trans. Vis. Comput. Graph.*, vol. 25, no. 1, pp. 299–309, Jan. 2019.
[166] A. Grau, M. Indri, L. Lo Bello, and T. Sauter, "Robots in industry: The past, present, and future of a growing collaboration with humans," *IEEE Ind. Electron. Mag.*, vol. 15, no. 1, pp. 50–61, Mar. 021.
[167] S. P. Ramu et al., "Federated learning enabled digital twins for smart cities: Concepts, recent advances, and future directions," *Sustain. Cities Soc.*, vol. 79, Apr. 2022, Art. no. 103663.
[168] M. I. Ali, P. Patel, J. G. Breslin, R. Harik, and A. Sheth, "Cognitive digital twins for smart manufacturing," *IEEE Intell. Syst.*, vol. 36, no. 2, pp. 96–100, Mar./Apr. 2021.
[169] P. K. R. Maddikunta et al., "Industry 5.0: A survey on enabling technologies and potential applications," *J. Ind. Inf. Integr.*, vol. 26, Mar. 2022, Art. no. 100257.
[170] K. A. Demir, G. Döven, and B. Sezen, "Industry 5.0 and human-robot co-working," *Procedia Comput. Sci.*, vol. 158, pp. 688–695, Oct. 2019.
[171] T. Brito, J. Queiroz, L. Piardi, L. A. Fernandes, J. Lima, and P. Leitão, "A machine learning approach for collaborative robot smart manufacturing inspection for quality control systems," *Procedia Manuf.*, vol. 51, pp. 11–18, Nov. 2020.

[172] M. H. Sayour, S. E. Kozhaya, and S. S. Saab, "Autonomous robotic manipulation: Real-time, deep-learning approach for grasping of unknown objects," *J. Robot.*, vol. 2022, Jun. 2022, Art. no. 2585656.
[173] Y. Sun, W. Wang, Y. Chen, and Y. Jia, "Learn how to assist humans through human teaching and robot learning in human–robot collaborative assembly," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 52, no. 2, pp. 728–738, Feb. 2022.
[174] T. Yu, J. Huang, and Q. Chang, "Mastering the working sequence in human-robot collaborative assembly based on reinforcement learning," *IEEE Access*, vol. 8, pp. 163868–163877, 2020.
[175] C. Gehrmann and M. Gunnarsson, "A digital twin based industrial automation and control system security architecture," *IEEE Trans. Ind. Informat.*, vol. 16, no. 1, pp. 669–680, Jan. 2020.
[176] T. R. Wanasinghe et al., "Digital twin for the oil and gas industry: Overview, research trends, opportunities, and challenges," *IEEE Access*, vol. 8, pp. 104175–104197, 2020.
[177] H.-Y. Chen and C.-H. Lee, "Vibration signals analysis by explainable artificial intelligence (XAI) approach: Application on bearing faults diagnosis," *IEEE Access*, vol. 8, pp. 134246–134256, 2020.
[178] S. Srinivasan, P. Arjunan, B. Jin, A. L. Sangiovanni-Vincentelli, Z. Sultan, and K. Poolla, "Explainable AI for chiller fault-detection systems: Gaining human trust," *Computer*, vol. 54, no. 10, pp. 60–68, Oct. 2021.
[179] V. Piroumian, "Digital twins: Universal interoperability for the digital age," *Computer*, vol. 54, no. 1, pp. 61–69, Jan. 2021.
[180] C. Hwang and T. Lee, "E-SFD: Explainable sensor fault detection in the ICS anomaly detection system," *IEEE Access*, vol. 9, pp. 140470–140486, 2021.
[181] X. Gao, R. Gong, Y. Zhao, S. Wang, T. Shu, and S.-C. Zhu, "Joint mind modeling for explanation generation in complex human-robot collaborative tasks," in *Proc. 29th IEEE Int. Conf. Robot Human Interact. Commun. (RO-MAN)*, 2020, pp. 1119–1126.
[182] E. Daglarli, "Computational modeling of prefrontal cortex for meta-cognition of a humanoid robot," *IEEE Access*, vol. 8, pp. 98491–98507, 2020.
[183] J. He, K. Yang, and H.-H. Chen, "6G cellular networks and connected autonomous vehicles," *IEEE Netw.*, vol. 35, no. 4, pp. 255–261, Jul./Aug. 2021.
[184] X. Chen, S. Leng, J. He, and L. Zhou, "Deep-learning-based intelligent intervehicle distance control for 6G-enabled cooperative autonomous driving," *IEEE Internet Things J.*, vol. 8, no. 20, pp. 15180–15190, Oct. 2021.
[185] A. Al-Dulaimi and X. Lin, "Reshaping autonomous driving for the 6G era," *IEEE Commun. Stand. Mag.*, vol. 4, no. 1, p. 10, Mar. 2020.
[186] K. Raats, V. Fors, and S. Pink, "Trusting autonomous vehicles: An interdisciplinary approach," *Transp. Res. Interdiscip. Perspect.*, vol. 7, Sep. 2020, Art. no. 100201.
[187] A. Basavaraju, J. Du, F. Zhou, and J. Ji, "A machine learning approach to road surface anomaly assessment using smart-phone sensors," *IEEE Sensors J.*, vol. 20, no. 5, pp. 2635–2647, Mar. 2020.
[188] Y. Fu, C. Li, F. R. Yu, T. H. Luan, and Y. Zhang, "A decision-making strategy for vehicle autonomous braking in emergency via deep reinforcement learning," *IEEE Trans. Veh. Technol.*, vol. 69, no. 6, pp. 5876–5888, Jun. 2020.
[189] S. Mozaffari, O. Y. Al-Jarrah, M. Dianati, P. Jennings, and A. Mouzakitis, "Deep learning-based vehicle behavior prediction for autonomous driving applications: A review," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 1, pp. 33–47, Jan. 2022.
[190] P. Luong, F. Gagnon, L.-N. Tran, and F. Labeau, "Deep reinforcement learning based resource allocation in cooperative UAV-assisted wireless networks," *IEEE Trans. Wireless Commun.*, vol. 20, no. 11, pp. 7610–7625, Nov. 2021.
[191] M. Samir, C. Assi, S. Sharafeddine, D. Ebrahimi, and A. Ghrayeb, "Age of information aware trajectory planning of UAVs in intelligent transportation systems: A deep learning approach," *IEEE Trans. Veh. Technol.*, vol. 69, no. 11, pp. 12382–12395, Nov. 2020.
[192] L. Sun, J. Liang, C. Zhang, D. Wu, and Y. Zhang, "Meta-transfer metric learning for time series classification in 6G-supported intelligent transportation systems," *IEEE Trans. Intell. Transp. Syst.*, early access, Mar. 6, 2023, doi: 10.1109/TITS.2023.3250962.
[193] S. Xiao, X. Ge, Q.-L. Han, and Y. Zhang, "Dynamic event-triggered platooning control of automated vehicles under random communication topologies and various spacing policies," *IEEE Trans. Cybern.*, vol. 52, no. 11, pp. 11477–11490, Nov. 2021.
[194] H. Wu, F. Lyu, C. Zhou, J. Chen, L. Wang, and X. Shen, "Optimal UAV caching and trajectory in aerial-assisted vehicular networks: A learning-based approach," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 12, pp. 2783–2797, Dec. 2020.
[195] E. Soares, P. Angelov, D. Filev, B. Costa, M. Castro, and S. Nageshrao, "Explainable density-based approach for self-driving actions classification," in *Proc. 18th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, 2019, pp. 469–474.
[196] B. M. Keneni et al., "Evolving rule-based explainable artificial intelligence for unmanned aerial vehicles," *IEEE Access*, vol. 7, pp. 17001–17016, 2019.
[197] D. Hamilton, K. Kornegay, and L. Watkins, "Autonomous navigation assurance with explainable AI and security monitoring," in *Proc. IEEE Appl. Imagery Pattern Recognit. Workshop (AIPR)*, 2020, pp. 1–7.
[198] M. Khonji, J. Dias, R. Alyassi, F. Almaskari, and L. Seneviratne, "A risk-aware architecture for autonomous vehicle operation under uncertainty," in *Proc. Proc. IEEE Int. Symp. Saf., Security, Rescue Robot. (SSRR)*, 2020, pp. 311–317.
[199] Y. Liu, X. Yang, W. Wen, and M. Xia, "Smarter grid in the 5G era: Integrating power Internet of Things with cyber physical system," *Front. Commun. Netw.*, vol. 2, p. 23, Jun. 2021.
[200] C. Zhong, H. Li, Y. Zhou, Y. Lv, J. Chen, and Y. Li, "Virtual synchronous generator of PV generation without energy storage for frequency support in autonomous microgrid," *Int. J. Electr. Power Energy Syst.*, vol. 134, Jan. 2022, Art. no. 107343.
[201] T. Dragičević, P. Siano, S. R. S. Prabaharan, and S. S. Reka, "Future generation 5G wireless networks for smart grid: A comprehensive review," *Energies*, vol. 12, no. 11, p. 2140, 2019.
[202] Y. Shakrina and H. Margossian, "A Stackelberg game-inspired model of real-time economic dispatch with demand response," *Int. Trans. Electr. Energy Syst.*, vol. 31, no. 11, 2021, Art. no. e13076.
[203] S. Bhattacharya et al., "Incentive mechanisms for smart grid: State of the art, challenges, open issues, future directions," *Big Data Cogn. Comput.*, vol. 6, no. 2, p. 47, 2022.
[204] P. K. R. Maddikunta et al., "Incentive techniques for the Internet of Things: A survey," *J. Netw. Comput. Appl.*, vol. 206, Oct. 2022, Art. no. 103464.
[205] C. Yapa, C. de Alwis, M. Liyanage, and J. Ekanayake, "Survey on blockchain for future smart grids: Technical aspects, applications, integration challenges and future research," *Energy Rep.*, vol. 7, pp. 6530–6564, Nov. 2021.
[206] H. Hui, Y. Ding, Q. Shi, F. Li, Y. Song, and J. Yan, "5G network-based Internet of Things for demand response in smart grid: A survey on application potential," *Appl. Energy*, vol. 257, Jan. 2020, Art. no. 113972.
[207] M. Tariq, M. Ali, F. Naeem, and H. V. Poor, "Vulnerability assessment of 6G-enabled smart grid cyber–physical systems," *IEEE Internet Things J.*, vol. 8, no. 7, pp. 5468–5475, Apr. 2021.
[208] S. Borenius, H. Hämmäinen, M. Lehtonen, and P. Ahokangas, "Smart grid evolution and mobile communications—Scenarios on the finnish power grid," *Electr. Power Syst. Res.*, vol. 199, Oct. 2021, Art. no. 107367.
[209] A. Sundararajan, A. S. Hernandez, and A. I. Sarwat, "Adapting big data standards, maturity models to smart grid distributed generation: Critical review," *IET Smart Grid*, vol. 3, no. 4, pp. 508–519, 2020.
[210] M. Alazab, S. Khan, S. S. R. Krishnan, Q.-V. Pham, M. P. K. Reddy, and T. R. Gadekallu, "A multidirectional LSTM model for predicting the stability of a smart grid," *IEEE Access*, vol. 8, pp. 85454–85463, 2020.
[211] M. Babar, M. U. Tariq, and M. A. Jan, "Secure and resilient demand side management engine using machine learning for IoT-enabled smart grid," *Sustain. Cities Soc.*, vol. 62, Nov. 2020, Art. no. 102370.
[212] S. M. Qaisar, F. Alsharif, A. Subasi, and A. Bensenouci, "Appliance identification based on smart meter data and event-driven processing in the 5G framework," *Procedia Comput. Sci.*, vol. 182, pp. 103–108, Mar. 2021.

[213] Y. Shen, W. Fang, F. Ye, and M. Kadoch, "EV charging behavior analysis using hybrid intelligence for 5G smart grid," *Electronics*, vol. 9, no. 1, p. 80, 2020.
[214] D. Sun et al., "Integrated human-machine intelligence for EV charging prediction in 5G smart grid," *EURASIP J. Wireless Commun. Netw.*, vol. 2020, no. 1, pp. 1–15, 2020.
[215] M. Kuzlu, U. Cali, V. Sharma, and Ö. Güler, "Gaining insight into solar photovoltaic power generation forecasting utilizing explainable artificial intelligence tools," *IEEE Access*, vol. 8, pp. 187814–187823, 2020.
[216] K. Zhang, J. Zhang, P.-D. Xu, T. Gao, and D. W. Gao, "Explainable AI in deep reinforcement learning models for power system emergency control," *IEEE Trans. Comput. Social Syst.*, vol. 9, no. 2, pp. 419–427, Apr. 2022.
[217] A. K. Singh, R. Singh, and B. C. Pal, "Stability analysis of networked control in smart grids," *IEEE Trans. Smart Grid*, vol. 6, no. 1, pp. 381–390, Jan. 2015.
[218] B. S. England and A. T. Alouani, "Real time voltage stability prediction of smart grid areas using smart meters data and improved Thevenin estimates," *Int. J. Electr. Power Energy Syst.*, vol. 122, Nov. 2020, Art. no. 106189.
[219] A. K. Bashir et al., "Comparative analysis of machine learning algorithms for prediction of smart grid stability," *Int. Trans. Elect. Energy Syst.*, vol. 31, no. 9, 2021, Art. no. e12706.
[220] D. Syed, H. Abu-Rub, S. S. Refaat, and L. Xie, "Detection of energy theft in smart grids using electricity consumption patterns," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, 2020, pp. 4059–4064.
[221] Z. Yan and H. Wen, "Electricity theft detection base on extreme gradient boosting in AMI," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–9, 2021.
[222] M. Ismail, M. F. Shaaban, M. Naidu, and E. Serpedin, "Deep learning detection of electricity theft cyber-attacks in renewable distributed generation," *IEEE Trans. Smart Grid*, vol. 11, no. 4, pp. 3428–3437, Jul. 2020.
[223] C. Hu, J. Yan, and X. Liu, "Adaptive feature boosting of multi-sourced deep autoencoders for smart grid intrusion detection," in *Proc. IEEE Power Energy Soc. Gen. Meeting (PESGM)*, 2020, pp. 1–5.
[224] S. Alizadehsalehi, A. Hadavi, and J. C. Huang, "From BIM to extended reality in AEC industry," *Autom. Construct.*, vol. 116, Aug. 2020, Art. no. 103254.
[225] S. Bilbow, "Developing multisensory augmented reality as a medium for computational artists," in *Proc. 15th Int. Conf. Tangible, Embedded, Embodied Interact.*, 2021, pp. 1–7.
[226] T. Andrade and D. Bastos, "Extended reality in IoT scenarios: Concepts, applications and future trends," in *Proc. 5th Exp. Int. Conf.*, 2019, pp. 107–112.
[227] A. Yastrebova, R. Kirichek, Y. Koucheryavy, A. Borodin, and A. Koucheryavy, "Future networks 2030: Architecture & requirements," in *Proc. 10th Int. Congr. Ultra Modern Telecommun. Control Syst. Workshops (ICUMT)*, 2018, pp. 1–8.
[228] R. K. D. Anjos et al., "Adventures in hologram space: Exploring the design space of eye-to-eye volumetric telepresence," in *Proc. 25th ACM Symp. Virtual Real. Softw. Technol.*, 2019, pp. 1–5.
[229] A. Clemm, M. T. Vega, H. K. Ravuri, T. Wauters, and F. De Turck, "Toward truly immersive holographic-type communication: Challenges and solutions," *IEEE Commun. Mag.*, vol. 58, no. 1, pp. 93–99, Jan. 2020.
[230] T. Huynh-The et al., "Blockchain for the metaverse: A review," *Future Gener. Comput. Syst.*, vol. 143, pp. 401–419, Jun. 2023.
[231] R. Chengoden et al., "Metaverse for healthcare: A survey on potential applications, challenges and future directions," *IEEE Access*, vol. 11, pp. 12765–12795, 2023.
[232] B. Siniarski et al., "Need of 6G for the metaverse realization," 2022, *arXiv:2301.03386*.
[233] A. K. Bashir et al., "A survey on federated learning for the healthcare metaverse: Concepts, applications, challenges, and future directions," 2023, *arXiv:2304.00524*.
[234] E. C. Strinati and S. Barbarossa, "6G networks: Beyond Shannon towards semantic and goal-oriented communications," *Comput. Netw.*, vol. 190, May 2021, Art. no. 107930.
[235] A. L. Imoize, O. Adedeji, N. Tandiya, and S. Shetty, "6G enabled smart infrastructure for sustainable society: Opportunities, challenges, and research Roadmap," *Sensors*, vol. 21, no. 5, p. 1709, 2021.
[236] M. J. Piran and D. Y. Suh, "Learning-driven wireless communications, towards 6G," in *Proc. Int. Conf. Comput., Electron. Commun. Eng. (iCCECE)*, 2019, pp. 219–224.
[237] "The power of AI combined with AR/VR technology." Accessed: Aug. 9, 2021. [Online]. Available: https://appen.com/solutions/ar-vr/
[238] K. Židek, P. Lazorík, J. Pitel', and A. Hošovskỳ, "An automated training of deep learning networks by 3D virtual models for object recognition," *Symmetry*, vol. 11, no. 4, p. 496, 2019.
[239] A. Manni, D. Oriti, A. Sanna, F. De Pace, and F. Manuri, "Snap2cad: 3D indoor environment reconstruction for AR/VR applications using a smartphone device," *Comput. Graph.*, vol. 100, pp. 116–124, Nov. 2021.
[240] G. Can, D. Mantegazza, G. Abbate, S. Chappuis, and A. Giusti, "Semantic segmentation on Swiss3DCities: A benchmark study on aerial photogrammetric 3D pointcloud dataset," *Pattern Recognit. Lett.*, vol. 150, pp. 108–114, Oct. 2021.
[241] M. Kim, K.-B. Park, S. H. Choi, J. Y. Lee, and D. Y. Kim, "AR/VR-based live manual for user-centric smart factory services," in *Proc. IFIP Int. Conf. Adv. Prod. Manage. Syst.*, 2018, pp. 417–421.
[242] N. Vretos et al., "Exploiting sensing devices availability in AR/VR deployments to foster engagement," *Virtual Real.*, vol. 23, no. 4, pp. 399–410, 2019.
[243] C.-J. Wu et al., "Machine learning at Facebook: Understanding inference at the edge," in *Proc. IEEE Int. Symp. High Perform. Comput. Archit. (HPCA)*, 2019, pp. 331–344.
[244] R. Sicat et al., "DXR: A toolkit for building immersive data visualizations," *IEEE Trans. Vis. Comput. Graph.*, vol. 25, no. 1, pp. 715–725, Jan. 2019.
[245] M. S. Mast, E. P. Kleinlogel, B. Tur, and M. Bachmann, "The future of interpersonal skills development: Immersive virtual reality training with virtual humans," *Human Resource Develop. Quart.*, vol. 29, no. 2, pp. 125–141, 2018.
[246] G. McLean and K. Osei-Frimpong, "Hey Alexa... Examine the variables influencing the use of artificial intelligent in-home voice assistants," *Comput. Human Behav.*, vol. 99, pp. 28–37, May 2019.
[247] C. J. Turner and W. Garn, "Next generation DES simulation: A research agenda for human centric manufacturing systems," *J. Ind. Inf. Integr.*, vol. 28, Jul. 2022, Art. no. 100354.
[248] B. Hayes and M. Moniz, "Trustworthy human-Centered automation through explainable AI and high-fidelity simulation," in *Proc. Int. Conf. Appl. Human Factors Ergonom.*, 2020, pp. 3–9.
[249] G. V. Pereira, P. Parycek, E. Falco, and R. Kleinhans, "Smart governance in the context of smart cities: A literature review," *Inf. Polity*, vol. 23, no. 2, pp. 143–162, 2018.
[250] A. M. Abdellatif, "Good governance and its relationship to democracy and economic development," in *Proc. Global Forum III Fight. Corruption Safeguarding Integr.*, vol. 20, 2003, p. 31.
[251] A. Mani and S. Mukand, "Democracy, visibility and public good provision," *J. Develop. Econ.*, vol. 83, no. 2, pp. 506–529, 2007.
[252] S. Pillay, "Corruption–The challenge to good governance: A South African perspective," *Int. J. Public Sector Manag.*, vol. 17, no. 7, pp. 586–605, 2004.
[253] S. K. Rao and R. Prasad, "Impact of 5G technologies on smart city implementation," *Wireless Pers. Commun.*, vol. 100, no. 1, pp. 161–176, 2018.
[254] D. Luckey, H. Fritz, D. Legatiuk, K. Dragos, and K. Smarsly, "Artificial intelligence techniques for smart city applications," in *Proc. Int. Conf. Comput. Civil Build. Eng.*, 2020, pp. 3–15.
[255] R. Brauneis and E. P. Goodman, "Algorithmic transparency for the smart city," *Yale J. Law Technol.*, vol. 20, p. 103, 2018.
[256] D. Thakker, B. K. Mishra, A. Abdullatif, S. Mazumdar, and S. Simpson, "Explainable artificial intelligence for developing smart cities solutions," *Smart Cities*, vol. 3, no. 4, pp. 1353–1382, 2020.
[257] M. A. Qureshi, L. Miralles-Pechuán, J. Payne, R. O'Malley, and B. Mac Namee, "Valve health identification using sensors and machine learning methods," in *IoT Streams for Data-Driven Predictive Maintenance and IoT, Edge, and Mobile for Embedded Machine Learning*. Cham, Switzerland: Springer, 2020, pp. 45–60.

[258] Q. Liu, H. Yuan, R. Hamzaoui, H. Su, J. Hou, and H. Yang, "Reduced reference perceptual quality model with application to rate control for video-based point cloud compression," *IEEE Trans. Image Process.*, vol. 30, pp. 6623–6636, 2021.

[259] R. Matheus, M. Janssen, and D. Maheshwari, "Data science empowering the public: Data-driven dashboards for transparent and accountable decision-making in smart cities," *Govern. Inf. Quart.*, vol. 37, no. 3, 2020, Art. no. 101284.

[260] M. A. Qureshi, A. Younus, and D. Greene, "TwitterCracy: Exploratory monitoring of twitter streams for the 2016 U.S. presidential election cycle," in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Disc. Databases*, 2016, pp. 71–75.

[261] G. Poghosyan, M. A. Qureshi, and G. Ifrim, "Topy: Real-time story tracking via social tags," in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Disc. Databases*, 2016, pp. 45–49.

[262] S. S. Matin and B. Pradhan, "Earthquake-induced building-damage mapping using explainable AI (XAI)," *Sensors*, vol. 21, no. 13, p. 4489, 2021.

[263] B. Mahbooba, M. Timilsina, R. Sahal, and M. Serrano, "Explainable artificial intelligence (XAI) to enhance trust management in intrusion detection systems using decision tree model," *Complexity*, vol. 2021, Jan. 2021, Art. no. 6634811.

[264] A. Morichetta, P. Casas, and M. Mellia, "EXPLAIN-IT: Towards explainable AI for unsupervised network traffic analysis," in *Proc. 3rd ACM CoNEXT Workshop Big DAta, Mach. Learn. Artif. Intell. Data Commun. Netw.*, 2019, pp. 22–28.

[265] O. Ayoub, A. Bianco, D. Andreoletti, S. Troia, S. Giordano, and C. Rottondi, "On the application of explainable artificial intelligence to lightpath QoT estimation," in *Proc. Opt. Fiber Commun. Conf.*, 2022, pp. M3F–5.

[266] A. Terra, R. Inam, S. Baskaran, P. Batista, I. Burdick, and E. Fersman, "Explainability methods for identifying root-cause of SLA violation prediction in 5G network," in *Proc. IEEE Global Commun. Conf.*, 2020, pp. 1–7.

[267] Z. Lv and N. Kumar, "Software defined solutions for sensors in 6G/IoE," *Comput. Commun.*, vol. 153, pp. 42–47, Mar. 2020.

[268] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," 2017, *arXiv:1702.08608*.

[269] R. R. Hoffman, S. T. Mueller, G. Klein, and J. Litman, "Metrics for explainable AI: Challenges and prospects," 2018, *arXiv:1812.04608*.

[270] A. Holzinger, A. Carrington, and H. Müller, "Measuring the quality of explanations: The system causability scale (SCS)," *KI-Künstliche Intelligenz*, vol. 34, pp. 193–198, Jan. 2020.

[271] I. E. Kumar, S. Venkatasubramanian, C. Scheidegger, and S. Friedler, "Problems with Shapley-value-based explanations as feature importance measures," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 5491–5500.

[272] M. J. Gacto, R. Alcalá, and F. Herrera, "Interpretability of linguistic fuzzy rule-based systems: An overview of interpretability measures," *Inf. Sci.*, vol. 181, no. 20, pp. 4340–4360, 2011.

[273] D. Gunning, M. Stefik, J. Choi, T. Miller, S. Stumpf, and G.-Z. Yang, "XAI—Explainable artificial intelligence," *Sci. Robot.*, vol. 4, no. 37, 2019, Art. no. eaay7120.

[274] M. Bohanec and I. Bratko, "Trading accuracy for simplicity in decision trees," *Mach. Learn.*, vol. 15, no. 3, pp. 223–250, 1994.

[275] T. A. Plate, "Accuracy versus interpretability in flexible modeling: Implementing a tradeoff using gaussian process models," *Behaviormetrika*, vol. 26, no. 1, pp. 29–50, 1999.

[276] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nat. Mach. Intell.*, vol. 1, no. 5, pp. 206–215, 2019.

[277] C. Rudin, C. Chen, Z. Chen, H. Huang, L. Semenova, and C. Zhong, "Interpretable machine learning: Fundamental principles and 10 grand challenges," 2021, *arXiv:2103.11251*.

[278] G. Vilone and L. Longo, "Explainable artificial intelligence: A systematic review," 2020, *arXiv:2006.00093*.

[279] G. Ke et al., "LightGBM: A highly efficient gradient boosting decision tree," *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–9.

[280] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 2016, pp. 785–794.

[281] R. Mitchell, E. Frank, and G. Holmes, "GPUTreeShap: Massively parallel exact calculation of SHAP scores for tree ensembles," *PeerJ Comput. Sci.*, vol. 8, p. e880, Apr. 2022.

[282] P. W. Koh and P. Liang, "Understanding black-box predictions via influence functions," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 1885–1894.

[283] A. Kuppa and N.-A. Le-Khac, "Adversarial xai methods in cybersecurity," *IEEE Trans. Inf. forensics security*, vol. 16, pp. 4924–4938, 2021.

[284] X. Zhao, W. Zhang, X. Xiao, and B. Lim, "Exploiting explanations for model inversion attacks," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 682–692.

[285] F. M. N. de Araújo, "XAIPrivacy—XAI with differential privacy," M.S. thesis, Faculdade de Engenharia, Universidade do Porto, Porto, Portugal, 2023.

[286] S. Goethals, K. Sörensen, and D. Martens, "The privacy issue of counterfactual explanations: Explanation linkage attacks," 2022, *arXiv:2210.12051*.

[287] M. T. Keane and B. Smyth, "Good counterfactuals and where to find them: A case-based technique for generating counterfactuals for explainable AI (XAI)," in *Proc. 28th Int. Conf. Case-Based Reason. Res. Develop.*, 2020, pp. 163–178.

[288] T. Senevirathna et al., "A survey on XAI for beyond 5G security: Technical aspects, use cases, challenges and research directions," 2022, *arXiv:2204.12822*.

[289] I. Stoica et al., "A Berkeley view of systems challenges for AI," 2017, *arXiv:1712.05855*.

[290] M. Veerappa, M. Anneken, N. Burkart, and M. F. Huber, "Validation of XAI explanations for multivariate time series classification in the maritime domain," *J. Comput. Sci.*, vol. 58, Feb. 2022, Art. no. 101539.

[291] "C/AISC—Artificial intelligence standards committee." Accessed: Oct. 1, 2023. [Online]. Available: https://standards.ieee.org/project/2894.html

[292] "Standard for XAI—Explainable AI working group." Accessed: Oct. 1, 2023. [Online]. Available: https://standards.ieee.org/project/2976.html

[293] P. J. Phillips et al., "Four principles of explainable artificial intelligence (draft)," NIST, Gaithersburg, MD, USA, Rep. NISTIR 8312, 2020.

[294] R. Righi, S. Samoili, M. L. Cobo, M. V.-P. Baillet, M. Cardona, and G. De Prato, "The AI techno-economic complex system: Worldwide landscape, thematic subdomains and technological collaborations," *Telecommun. Policy*, vol. 44, no. 6, 2020, Art. no. 101943.

[295] G. De Prato et al., "The AI techno-economic segment analysis," Publications Office Eur. Union, Luxembourg city, Luxembourg, Joint Research Centre, Rep. EUR 29952 EN, 2019.

[296] G. T. Reddy et al., "Analysis of dimensionality reduction techniques on big data," *IEEE Access*, vol. 8, pp. 54776–54788, 2020.

[297] T. R. Gadekallu, Q.-V. Pham, T. Huynh-The, S. Bhattacharya, P. K. R. Maddikunta, and M. Liyanage, "Federated learning for big data: A survey on opportunities, applications, and future directions," 2021, *arXiv:2110.04160*.

[298] Y. Liu, J. James, J. Kang, D. Niyato, and S. Zhang, "Privacy-preserving traffic flow prediction: A federated learning approach," *IEEE Internet Things J.*, vol. 7, no. 8, pp. 7751–7763, Aug. 2020.

[299] J. Song, W. Wang, T. R. Gadekallu, J. Cao, and Y. Liu, "EPPDA: An efficient privacy-preserving data aggregation federated learning scheme," *IEEE Trans. Netw. Sci. Eng.*, vol. 10, no. 5, pp. 3047–3057, Sep./Oct. 2023.

[300] D. Gunning and D. W. Aha, "DARPA's explainable artificial intelligence program," *AI Mag*, vol. 40, no. 2, p. 44, 2019.

[301] M. Veale and F. Z. Borgesius, "Demystifying the draft EU artificial intelligence act—Analysing the good, the bad, and the unclear elements of the proposed approach," *Comput. Law Rev. Int.*, vol. 22, no. 4, pp. 97–112, 2021.

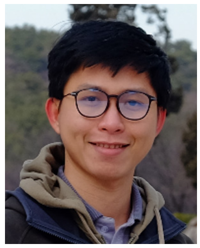
[302] S. Yarkoni et al., "Quantum shuttle: Traffic navigation with quantum computing," in *Proc. 1st ACM SIGSOFT Int. Workshop Archit. Paradigms Eng. Quantum Softw.*, 2020, pp. 22–30.

[303] D. D. F. Maesa and P. Mori, "Blockchain 3.0 applications survey," *J. Parallel Distrib. Comput.*, vol. 138, pp. 99–114, Apr. 2020.

SHEN WANG (Senior Member, IEEE) received the M.Eng. degree from Wuhan University, China, and the Ph.D. degree from Dublin City University, Ireland. He is currently an Assistant Professor with the School of Computer Science, University College Dublin, Ireland. He has been involved with several EU projects as a Co-PI, a WP, and a Task Leader in big trajectory data streaming for air traffic control and trustworthy AI for intelligent cybersecurity systems. Some key industry partners of his applied research are IBM Research Brazil, Boeing Research and Technology Europe, and Huawei Ireland Research Centre. His research interests include connected autonomous vehicles, explainable artificial intelligence, and security and privacy for mobile networks.

M. ATIF QURESHI received the B.S. degree in computer science from the University of Karachi, Pakistan, the M.S. degree in computer science from the Korea Advanced Institute of Science and Technology, South Korea, and the joint Ph.D. degree in computer science from the National University of Ireland Galway, Ireland, and the University of Milano-Bicocca, Italy. He is a Lecturer/Assistant Professor with Technological University Dublin, Ireland. His research interests include natural language processing, machine learning, disinformation space, and explainable artificial intelligence. He has contributed to various projects funded by Science Foundation Ireland and the Irish Research Council as a Principal Investigator and a Technical Lead for those funded by Enterprise Ireland and the EU calls and licensed an outcome to a leading media organization of Ireland in social media analytics. He is passionate about applied research focusing on the tight coupling of business needs and analytics to create value and impact.

LUIS MIRALLES-PECHUÁN received the bachelor's and Ph.D. degrees in computer science from the University of Murcia, Spain. He is currently a Lecturer with Technological University Dublin. He worked as a full-time Researcher/Lecturer with Panamerican University, Mexico, for more than three years. He started a Ph.D. in 2012 on creating new approaches within the Online Advertising world. During his Ph.D., he got familiar with ML and many papers on topics related to how to apply ML to online advertising. After finishing his Ph.D., he worked in postdoc levels I and II with CeADAR, University College Dublin, and there, he won the prize for the best student paper at the Digital Forensic Conference. His favorite topic is how to apply reinforcement learning to fight the COVID-19 pandemic and plan the containing levels considering both public health and the economy. Finally, he has expertise in human activity recognition and generalized zero-shot learning and applying machine learning to improve the accessibility of websites.



THIEN HUYNH-THE (Senior Member, IEEE) received the B.S. degree in electronics and telecommunication engineering and the M.Sc. degree in electronics engineering from the Ho Chi Minh City University of Technology and Education (HCMUTE), Vietnam, in 2011 and 2013, respectively, and the Ph.D. degree in computer science and engineering from Kyung Hee University (KHU), South Korea, in 2018. From March 2018 to August 2018, he was a Postdoctoral Researcher with the Ubiquitous Computing Laboratory, KHU. From September 2018 to May 2022, he was a Postdoctoral Researcher with the ICT Convergence Research Center, Kumoh National Institute of Technology, South Korea. He is currently a Lecturer with the Department of Computer and Communication Engineering, HCMUTE. His current research interests include digital image processing, radio signal processing, computer vision, wireless communications, IoT applications, machine learning, and deep learning. He was a recipient of the Superior Thesis Prize awarded by KHU and the Golden Globe Award 2020 for Vietnamese Young Scientists by Central Ho Chi Minh Communist Youth Union associated with the Ministry of Science and Technology.

THIPPA REDDY GADEKALLU (Senior Member, IEEE) received the bachelor's degree in computer science and engineering from Nagarjuna University, India, in 2003, the master's degree in computer science and engineering from Anna University, Chennai, India, and the Ph.D. degree in machine learning from the Vellore Institute of Technology (VIT), Vellore, India, in 2017. He is currently working as a Chief Engineer with Zhongda Group, Jiaxing, China, as well as an Associate Professor with the School of Information Technology and Engineering, VIT, the Division of Research and Development, Lovely Professional University, Phagwara, India, and the Center of Research Impact and Outcome, Chitkara University, India. He has more than 14 years of experience in teaching. He has published more than 200 international/national publications. His current areas of research include machine learning, Internet of Things, deep neural networks, blockchain, and computer vision.

MADHUSANKA LIYANAGE (Senior Member, IEEE) received the Doctor of Technology degree in communication engineering from the University of Oulu, Oulu, Finland, in 2016. From 2011 to 2012, he worked as a Research Scientist with the I3S Laboratory and Inria, Sophia Antipolis, France. He is currently an Assistant Professor/Ad Astra Fellow and the Director of Graduate Research with the School of Computer Science, University College Dublin, Ireland. He is also acting as an Adjunct Processor with the Center for Wireless Communications, University of Oulu and the Department of Electrical and Information Engineering, University of Ruhuna, Sri Lanka. From 2015 to 2018, he was a Visiting Research Fellow with CSIRO, Australia; the Infolabs21, Lancaster University, U.K.; the Department of Computer Science and Engineering, The University of New South Wales, Australia; the School of IT, University of Sydney, Australia; LIP6, Sorbonne University, France; and the Department of Computer Science and Engineering, The University of Oxford, U.K. He is an expert consultant at the European Union Agency for Cybersecurity (ENISA). In 2021, he was elevated as a Funded Investigator of the Science Foundation Ireland CONNECT Research Centre, Ireland. His research interests are 5G/6G, SDN, IoT, blockchain, MEC, mobile, and virtual network security. In 2020, he received the "2020 IEEE ComSoc Outstanding Young Researcher" Award from IEEE ComSoc EMEA. He was also a recipient of the prestigious Marie Skłodowska-Curie Actions Individual Fellowship from 2018 to 2020. He was ranked among the World's Top 2% Scientists (2020) in the list prepared by Elsevier BV, Stanford University, USA. He was also awarded an Irish Research Council (IRC) Research Ally Prize as part of the IRC Researcher of the Year 2021 awards for the positive impact he has made as a supervisor. Moreover, he is an expert reviewer at different funding agencies in France, Qatar, UAE, Sri Lanka, and Kazakhstan. More info: www.madhusanka.com.