Interpreting AI for Networking: Where We Are and Where We Are Going

Interpreting AI for Networking: Where We Are and Where We Go

Tianzhu Zhang, *Member, IEEE,* Han Qiu, Marco Mellia, *Fellow, IEEE,* Yuanjie Li, *Member, IEEE,* Hewu Li, and Ke Xu, *Senior Member, IEEE,*

***Abstract*—In recent years, Artificial Intelligence (AI) techniques have been increasingly adopted to tackle networking problems. Although AI algorithms can deliver high-quality solutions, most of them are inherently intricate and erratic for human cognition. This lack of interpretability tremendously hinders the commercial success of AI-based solutions in practice. To cope with this challenge, networking researchers explore eXplainable AI (XAI) techniques to make AI models interpretable, manageable, and trustworthy. In this paper, we overview the application of AI in networking and discuss the necessity for interpretability. Next, we review the current research on interpreting AI-based networking solutions and systems. At last, we envision future challenges and directions. The ultimate goal of this paper is to present a general guideline for AI and networking practitioners and motivate the continuous advancement of AI-based solutions in modern communication networks.**

***Index Terms*—XAI for Networking**

I. Introduction

Last decade has witnessed an unprecedented resurge of interest in Artificial Intelligence (AI) in industry and academia. Nowadays, AI-based solutions have been widely deployed across various sectors, including health care, business intelligence, and industrial manufacturing [1]. Meanwhile, with the rapid deployment of mobile networks, edge computing, Internet of Things (IoT), and Unmanned Aerial Vehicle (UAV), modern networking systems are becoming cumulatively diverse, ad hoc, and complex to manage. The fast emergent interactive applications and network services have assorted performance characteristics and depend upon fine-granular, passive and active traffic monitoring and real-time analytics for Quality-of-Experience (QoE) management. Consequently, network management has become an extremely daunting undertaking. Traditional network operators heavily lean on domain-specific knowledge to build rule-based procedures and heuristics, which become burdensome to sustain the same level of effectiveness upon network expansions or scenario changes. As a result, a plethora of research has been devoted to applying AI techniques to solve problems in heterogeneous modern networking systems, as illustrated by the scenarios in Fig. 1. Most of these AI-augmented solutions managed to attain superior performance compared to the traditional handcrafted, rule-based heuristic solutions [2]–[4].

However, performance improvement cannot directly map to the success of AI for networking. The current trend of using AI models, especially Deep Learning (DL) models, is to treat them as blackboxes. Their complexity keeps growing to include more parameters since complex DL models can better approximate universal functions, which leads to great success in solving famous computer vision problems. However, applying AI models to solve networking problems has many practical obstacles. (1) *Data discrepancy*: unlike image and text data, the networking data have inherent peculiarities such as time diversity, space diversity, and the abundance of categorical features. It is thus non-trivial to replicate the success of AI in networking due to both the lack of labeled data and the diversity of the scenarios. (2) *Feasibility*: although existing AI-based solutions mainly operate in the control plane, a recent trend pushes the AI frontiers to the data plane, which remains challenging given the scarce resources therein. (3) *Robustness*: there are many vulnerabilities for the current AI systems, which could let attackers manipulate the AI solutions and thus impact the network and QoE. (4) *Trust*: decisions made by sophisticated AI models usually entail a myriad of parameters and non-linear transformations that are too complex for humans to understand and to trust. This last point is especially essential in networks, where the operators need to understand the implications of a decision. Promoting the trust for AI-based solutions can realize the ultimate goal of responsible AI [5].

To overcome these issues, researchers work on eXplainable AI (XAI) to interpret the inference process of AI models. XAI can boost performance boosts with less complex model structures and fewer parameters. The robustness against adversarial attacks and the trustworthiness of the stakeholders can also be improved. However, very few works specifically concentrated on XAI for networking.

The purpose of this work is to fill this blank in two steps. First, we review the applications of AI techniques in the modern networking domain and discuss the current research endeavors for interpreting AI in networking. Second, we present the challenges and future perspectives. In summary, our goal is to provide a first-hand guideline on XAI for practitioners in the networking community and catalyze the sustainable development of AI in networking.

T. Zhang is with Nokia Bell Labs, Paris-Saclay, France. (Email: tianzhu.zhang@nokia-bell-labs.com)

H. Qiu (corresponding author), Y. Li, and H. Li are with Institute for Network Sciences and Cyberspace, BNRist, Tsinghua University, Beijing, China. (Email: {qiuhan, yuanjiel}@tsinghua.edu.cn, lihewu@cernet.edu.cn)

M. Mellia is with the Department of Control and Computer Engineering, Politecnico di Torino, Turin, Italy. (Email: marco.mellia@polito.it)

K. Xu is with the Department of Computer Science, BNRist, Tsinghua University, Beijing, China. (Email: xuke@tsinghua.edu.cn)

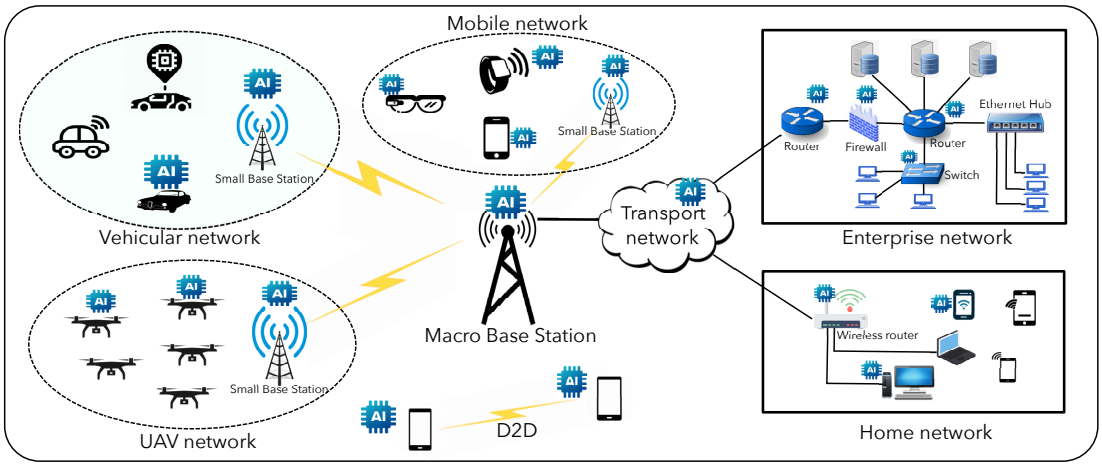

Fig. 1. An illustrative example of heterogeneous AI-based communication networks and systems, including vehicular network, mobile network, UAV network, enterprise network, home network, D2D communications, and transport network.

II. AI IN NETWORKING: A GENERAL OVERVIEW

In this section, we give a general overview of the motivations of AI-based solutions in networking. Then, we highlight the urgent need for interpretable AI-based solutions.

A. Benefits of AI in solving networking problems

Traditionally, network operators resort to rule-based and modeling-based algorithms and heuristics to address both in-network problems (e.g., packet routing, traffic classification) and end-to-end issues (e.g., congestion control, QoE prediction) [3]. However, with the growing scale and complexity of modern networks and the diverse requirements of applications, these approaches face severe limitations.

First, it is arduous for rule-based algorithms to comprehensively consider the related factors that can explicitly or implicitly impact the performance in a vast problem space. For instance, high-speed traffic processing stacks (e.g., FD.io VPP, Open vSwitch with DPDK) need to not only make the most suitable forwarding decisions but also consider miscellaneous low-level system details (e.g., buffer occupancy, cache locality, batch sizes) to optimally schedule resource and realize the intended network services at line rate. Second, the static algorithms cannot be improved by incremental learning, which makes them susceptible to recurrent execution pitfalls (e.g., load imbalance) and adversarial maneuvers (e.g., DDoS attacks). Third, the rules are primarily constructed based on human experience in specific scenarios and must be thoroughly adjusted upon domain or environment shift, making rule-based algorithms challenging to be adapted for reuse. For example, migrating a rule-based TCP congestion control algorithm from wired networks to wireless networks requires scrutinizing the additional impact of signal interference, link failures, and other performance impairments [4], which demands detailed knowledge on both networks.

As detailed in [3], AI techniques have been widely applied for various in-network and end-to-end tasks. We review the literature and summarize the commonly employed AI models and their use cases in Table I. Compared to traditional approaches, AI-based solutions possess several key advantages: (1) AI models can discover hidden patterns and automatically extract insights from voluminous data of heterogeneous sources, which makes them practical for analytics tasks in large-scale environments with abounding correlated factors (e.g., anomaly detection, root cause analysis). (2) AI techniques can efficiently capture and adapt to the temporal and spatial network dynamics. For instance, unlike traditional algorithms that identify network congestion through predefined static triggers, ML algorithms can proactively exploit various information to predict bottleneck conditions. (3) AI-based solutions can autonomously drive networks without human intervention, which is crucial to fulfilling the vision of zero-touch networks. Although the novel network softwarization technologies such as Software-Defined Networking (SDN) and Network Function Virtualization (NFV) significantly reduce operational costs, they still rely on static, hand-crafted algorithms for network service management and resource provisioning. AI techniques such as Reinforcement Learning (RL) can be integrated with existing frameworks to grant unprecedented flexibility and intelligence on network automation. (4) With the proliferation of transfer learning techniques, pre-trained AI models can be possibly refactored for networking tasks in different settings.

B. Commonly adopted AI models

AI models have quite different performance characteristics and interpretation overhead. Commonly adopted models are Naive Bayes (NB), Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), and Deep Neural Network (DNN) [6]. Based on internal functioning, these models can be categorized as *transparent AI models* and *opaque AI models*.

Transparent AI models: Transparent models are simple by design and can be readily presented to humans through simulation, decomposition, or algorithmic analysis [5]. AI models such as DT and NB are transparent and self-explanatory. For instance, DT consists of a hierarchy of nodes to split input

TABLE I
COMMONLY EMPLOYED AI MODELS TO SOLVE DIFFERENT NETWORKING PROBLEMS

AI model	Complexity	Transparency	Common use cases
Decision Tree (DT)	Low	High	Unsupervised traffic classification, misuse intrusion detection
Naive Bayes (NB)	Low	High	Supervised traffic classification
Random Forest (RF)	High	Low	QoE prediction, routing
Support Vector Machine (SVM)	High	Low	Supervised traffic classification, congestion control
Deep Neural Network (DNN)	High	Low	Traffic prediction, routing, congestion control, resource scheduling, anomaly detection, QoE prediction

data and leaves to represent predictions. In the networking domain, DTs are usually employed to tackle scenarios that are fault-tolerant and time-critical. NB is based on the assumption that the input features are independent of each other. If the "naive" assumption holds, NB can make precise predictions with relatively few training data. Both models offer means to understand their decision-making process.

Opaque AI models: AI models whose predictions cannot be easily communicated are deemed opaque. RF, SVM, and NN are typical opaque models. RF is an ensemble learning method that combines the predictions of multiple DTs to improve accuracy. SVM represents input data as points in a multi-dimensional space and uses hyperplanes to separate them into classes. DNNs are inspired by the structure of biological neurons in human brains. Sophisticated Deep Neural Networks (DNNs) can contain millions or even billions of parameters and are widely used for various complex tasks where proved to have outstanding performance.

Existing research shows that compared to transparent AI models, opaque AI models are less interpretable but much more proficient in capturing non-linear patterns and solving complex tasks. For instance, a linear regression model is intuitive to explain as the linear relationship automatically provides a straightforward mapping between feature input and target output. However, linear regression oversimplifies the context and often fails to deal with complex real-world problems. Similarly, the inference of a DT can be handily simulated by humans. Nonetheless, DTs suffer from over-fitting and are non-trivial to generalize. Researchers usually resort to ensemble methods like RF, which results in more accurate prediction but are innately too equivocal to interpret. Another significant difference between transparent and opaque AI models is the resource requirement. Transparent models have much simpler and fewer operations compared to opaque models whose scale can be prohibitively large. For instance, a powerful DNN model can contain billions of parameters that require specialized hardware (e.g., GPUs or TPUs) to accommodate the enormous computation and memory cost. Nevertheless, it is impractical to expect network devices such as programmable switches, routers, or smartNICs to spare adequate resources to deploy and serve these high-performance but heavyweight models.

C. The need for explainable AI in networking

Although some AI-based networking solutions still adopt transparent models, they are not the majority in the current research trend. According to a recent survey [6], most of the existing AI-based networking solutions are built on opaque models, which considerably plateau the development of AI-based networking solutions. Compared to well-established AI application domains such as computer vision, networking tasks have disparate time and spatial diversities and abundant categorical features (e.g., IP addresses, ports, paths). These tasks call for the availability of labeled data which is unfortunately hard to obtain. The continuing moving targets such as new applications, protocols, and patterns only make the situation more complex. Naively applying these opaque AI models without interpretation raises concerns about their robustness, reliability, and trustworthiness. Also, networking tasks customarily have a high reliance on domain-specific knowledge and experience, and human experts will always play an irreplaceable role [7]. As networks are destined to become more intelligent in the future, it is beneficial to consolidate the abilities of human experts and AI models to deliver the most performant and cost-efficient solutions. However, the opaqueness of most AI models completely blocks human involvement. XAI techniques can explain the inner workings of the AI-based solutions in understandable formats to let network/AI experts inspect and dissect the current solutions and craft high-level augmentations with domain expertise. Specifically, XAI techniques can enhance AI-based networking solutions in the following four aspects.

1) Performance: Albeit AI-based solutions can make satisfactory predictions and decisions, the underlying AI models are not immune to undesirable results or errors. An error can still occur in any stage of a model development cycle due to mislabelled data, poor feature selection, model drift, or deficient design. XAI techniques provide means to scrutinize the model and reveal potential bias and variances. AI experts can subsequently discern whether a particular network policy made by a model is derived from the intended portion of input data or control logic and take the correct measures to make the model more generalizable to network and system dynamics, e.g., adjusting the dataset, changing the feature set, tuning the hyper-parameters, redesigning the model architecture. In addition, network engineers can capitalize on the generated interpretations to pinpoint the decisive factors for a given AI-based solution and perform tailored optimizations based on the high-level observations of the problem settings.

2) Feasibility: Besides performance improvement, XAI can assist in model refinement. Considering the complexity of many existing opaque AI models, it is challenging to accommodate them using resource-constrained networking devices. Researchers are exploring different methods to reduce the AI models and fit them into small devices, commonly referred to as "Tiny AI". As shown in [8], XAI can be combined with these methods to expose the redundant operations and

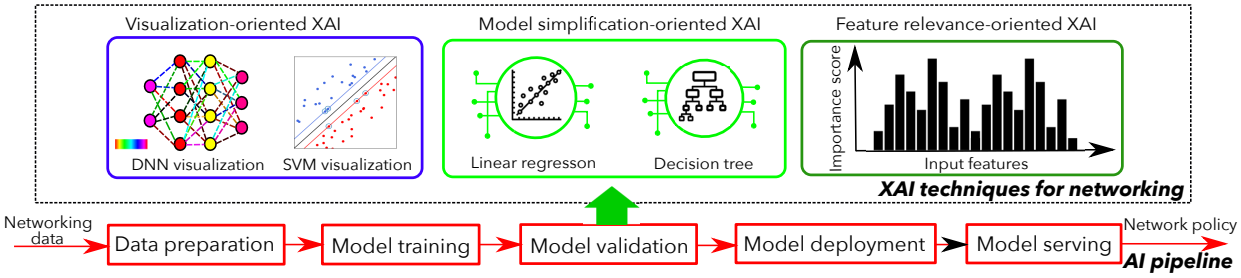

Fig. 2. A summary of the existing XAI techniques, i.e., visualization-oriented, model simplification-oriented, and feature analysis-oriented XAI.

features and shrink the incurred computation cost, processing latency, memory footprint, and energy consumption of existing DL-based networking systems. More advanced XAI methods are expected to progressively make AI models versatile and suitable for networking problems in the future.

3) Robustness: AI models, especially the DNN models, are well known to be non-robust against adversarial attacks. Using AI-based solutions in networking will inherit this vulnerability, which will threaten the models' applicabilities and even the security of the resulting networking systems. For instance, researchers have introduced DNN models in intrusion detection systems and achieved better detection accuracy than traditional approaches. However, attackers can introduce malicious modifications (e.g., several bits in a network packet) to generate Adversarial Examples (AE) that can easily mislead the DNN-based detectors. XAI techniques can help defenders understand their vulnerabilities from both the DNN model and networking data aspects. Besides the security against the famous adversarial attacks, XAI can also assist network administrators to discover the (otherwise hidden) security threats and loopholes in an interpretable way.

4) Trust: Humans are naturally reluctant to trust non-justifiable decisions made by AI-based solutions without proper insights into the internal inference mechanisms [9]. XAI is a fundamental requirement for solutions based on opaque models [10]. Depending on the target users, even solutions based on transparent AI models might still need to be explained. For instance, although transparent models such as DT and NB are relatively intuitive for ML engineers to understand, they might not be accessible for users without technical experience. In these cases, XAI techniques provide straightforward and non-ambiguous interpretations to the involved audience with or without a proper technical background. This benefit is especially crucial for mission-critical networks, e.g., banking, satellite, UAV, and transportation system networks, where predictable policies and deterministic behaviors are highly valued. Thus, explaining the AI models can expedite the validation of functional coherence, constraints violations, ethical customs, and legal obligations and make their decisions and recommendations more trustworthy, accountable, and dependable to human users.

III. Current XAI-based Solutions in Networking

XAI methods can be categorized using different criteria. Depending on the interpretation scope, XAI methods can be either global or local. Global methods strive for comprehensive model interpretation while local methods provide interpretations on specific prediction instances. Based on the reliance on specific AI models, XAI methods can be model-agnostic or model-dependent. Model-dependent methods are custom-made for specific models, while model-agnostic methods are technically applicable to any AI model. In this section, we classify existing XAI research in networking based on interpretation techniques, namely *visualization*, *model simplification*, and *feature relevance analysis*, as shown in Fig. 2.

A. Visualization-oriented XAI

The most straightforward XAI method is explaining through visualization, which entails visual augmentation and (optionally) dimensionality reduction techniques to generate simple illustrations of an AI model's internal operations and interactions. Beliard et al. [11] proposed a platform to visualize the inference process of a commercial-grade network traffic classification engine based on Convolutional Neural Networks (CNNs). The platform can generate a set of graphs to illustrate the classification process and highlight the most salient features. Human users can thus develop a better understanding of CNN's classification process by interacting with the graphs.

B. Model simplification-oriented XAI

The model simplification method builds a functionally similar yet much-simplified model (e.g., linear models) to elucidate the inference process. For instance, Morichetta et al. [12] targeted the unsupervised traffic classification problem and trained an SVM-based classifier. Then, they integrated the Local Interpretable Model-agnostic Explanations (LIME) approach to explain the specific clustering results. LIME constructs an interpretable model coherent with the SVM model for a given prediction instance and perturbs the input to locate the most influential features. Similarly, Sun et al. [13] presented their preliminary research on wireless multi-channel power allocation. The authors leveraged Meijer G-function to represent a NN model and to render a low-dimensional explainable symbolic representation. As Meijer G-function has an ample search space, there is no guarantee that it is the most representative function for the NN model. Meng et al. [8] proposed two methods to interpret DL-based networking systems. They utilized teacher-student training to build DTs for local networking systems and hyper-graph formulations to

TABLE II
THE RELATED WORKS FOR XAI IN NETWORKING

Research item	Problem domain	XAI technique	Interpretation scope	Model-agnostic	Target model
Beliard et al. [11]	Traffic classification	Visualization	Local	No	DNNs
Morichetta et al. [12]	Video quality classification	Model simplification	Local	Yes	-
Sun et al. [13]	Wireless channel allocation	Model simplification	Local	No	DNNs
Meng et al. [8]	Interpreting DL-based system	Model simplification	Local	No	DNNs
Guo et al. [14]	Wireless service provisioning	Feature relevance analysis	Local	No	DDDQN
Terra et al. [15]	5G root cause identification	Feature relevance analysis	Global/Local	Yes	-

generate interpretable policies for global networking systems. The proposed methods were applied to interpret three real-world DL-based systems (i.e., video streaming, flow scheduling, and SDN-based routing) and presented more accurate interpretation results than LIME and LEMNA, another prevalent XAI method approximates a local region of the complex decision boundary with an interpretable model.

C. Feature relevance-oriented XAI

Feature relevance analysis methods compute a feature relevance score to assess each feature's impact on the final decision. For example, Guo et al. [14] proposed a DRL method for optimal service provisioning in the UAV-based wireless networks and conducted local feature analysis using a sample configuration to interpret and highlight the determinant features leading to specific predictions. Terra et al. [15] tackled the interpretability of an XGBoost model that predicts the latency violation in 5G networks. The authors evaluated several classical XAI methods and recommended using SHAP that rendered the most proper interpretation.

Although these endeavors to interpret AI-based solutions have borne some fruit, XAI is still in its infancy in the networking field. According to Table II, existing methods either heavily rely on state-of-the-art XAI techniques designed for general purposes or have limited interpretation scope only applicable to some specific AI models.

IV. CHALLENGES AND FUTURE PERSPECTIVES

With the deployment of 5G and the inception of 6G standardization, there is an urgent need for end-to-end network automation. Several initiatives were established to strive for AI-driven self-managed networks, e.g., ETSI's Zero-touch network and Service Management group. XAI is deemed a fundamental building block to bestow the next-generation networks with self-managing, self-healing, and self-optimizing capabilities. However, XAI still has many impediments to overcome to unleash its full potential for automated network management. This section addresses five fundamental perspectives of XAI, including the network-specialized interpretation, performance improvement, model refinement, robustness, and trust fostering.

A. Specialized XAI for networking problems

As shown in Sec. III, most existing works directly adopt state-of-the-art XAI methods such as LIME and SHAP which are not natively designed to exploit the unique characteristics of modern networking systems and can lead to inconsistent results. For instance, as shown in [15], due to the unique patterns of network data, LIME failed to produce consistent interpretations when multiple features have similar impacts on one prediction, which can cause undesirable consequences. It is thus essential to consider the peculiarities of the target problem and implement bespoke XAI methods compatible with the corresponding network and system settings. To this end, Meng et al. [8] pioneered the design of specialized XAI methods for DL-based networking systems. Despite the promising results, their methods cannot explain Recurrent Neural Network (RNN)-based systems, and the performance for more complex DNNs is still unexplored. With the ascending complexity of modern networks, more XAI methods designed for diversiform combos of AI models and network settings are expected to be implemented to provide interpretations for AI-based network services, applications, and systems.

B. XAI for performance improvement

Future XAI methods should generate more advanced interpretations to facilitate performance improvement. Current XAI methods only extract mappings between input features and output predictions, which are subsequently analyzed and extrapolated by human experts to uncover the decisive factors. XAI methods should produce advanced observations and straightforward suggestions for automatic performance optimization at both model and system levels. Specifically, at the model level, XAI methods should explicitly indicate the steps to improve the quality of predictions, e.g., fine-tune parameters, augment the collected data, or simplify the model. At the system level, XAI methods should pinpoint the most desirable execution configurations for the deployed AI-based solutions, such as the intended network environment (e.g., data centers vs. ISP networks), traffic types, and model serving schemes. In some cases, XAI should enable acceptable tradeoffs between different performance metrics such as accuracy, latency, and energy cost. For instance, a DNN can be partitioned for collaborative training and inference based on real-time network dynamics. Its responsiveness can also be enhanced by adding multiple side branches with different degrees of accuracy (e.g., early-exit). Thus, future XAI methods need to integrate other cutting-edge analytic tools to extract insights and associate actions with decisions productively.

C. XAI for feasibility-oriented model refinement

Traditional network management runs in the control plane to react to the network events within milliseconds which can

only make decisions based on a few data and fail to capture more fine-granular statistics. With the proliferation of AI in networking, it is essential to leverage the abundant traffic features in the data plane to build more cognitive solutions for network management tasks such as traffic classification, congestion control, and QoE management. Given the resource constraints of network devices and the over-parameterization of many AI models, it is necessary to distill the most relevant features and reduce the model complexity to fit AI-based solutions into production-grade data planes. Although XAI can be used for model refinement, few prior research explicitly addresses this issue. Future XAI methods should pinpoint the most suitable model refinement strategy for different network and system settings. For instance, given the available capacities of a NetFPGA, a DNN's computation and memory footprints can be shrunk using model compression techniques such as pruning, quantization, and knowledge distillation.

D. The robustness of XAI

Another critical challenge for AI-driven network management is the robustness against malicious attacks. Although XAI can enhance the robustness of the AI-based solutions by exposing the vulnerabilities therein, XAI methods themselves are also susceptible to adversarial attacks. By purposefully manipulating the input data, existing XAI methods (e.g., LIME) can be misled to produce unreliable or irrelevant explanations. The unique characteristics of networking problems further introduce a different dimension to this challenge. Therefore, to guarantee unbiased interpretations for AI-based networking solutions, it is necessary to propose reliable benchmarks that can comprehensively assess the consistency, correctness, and scalability of the XAI methods. Besides, it is equally important to defend the XAI methods against adversarial attacks. The defense can be based on mechanisms designed to detect and prevent malicious attacks. Proactive defense schemes, such as shield execution and traffic encryption, are also viable options.

E. XAI for trust fostering

Most existing XAI methods are still evaluated in simulated or controlled environments, and their performance cannot sufficiently reflect real-world circumstances. This reality gap immensely impedes the acceptance of AI-based solutions, especially for the envisaged 6G networks where many mission-critical services are expected to be managed [9]. Unfortunately, existing XAI cannot be seamlessly integrated into network systems to interpret models on the fly. To further promote the trust of AI across the networking community, more system-level supports, such as standard APIs and software development kits, are needed to fuse XAI methods into the production network environment and facilitate the real-time, automatic inspection and validation of different AI-based solutions. By continuously providing high-quality inference with justified interpretations, the network operators and other stakeholders will become more accustomed to the AI-based solutions and more inclined to trust their decisions.

V. Conclusion

Despite the unprecedented success of AI techniques, most AI-based solutions are built on non-transparent models that are hard to interpret. Although XAI techniques keep gaining momentum, little attention has been paid to their applications in modern networking systems. In this paper, we gave a general overview of XAI in networking. We specifically reviewed the current status of AI in networking and discussed the motivations for XAI. We also reviewed existing XAI research that interprets AI-based solutions and discussed future challenges. Although XAI in networking is far from maturity, this paper can serve as primitive guidance for the incremental melioration of AI-based networking solutions.

Acknowledgement

This work is supported by the Natural Science Foundation of China under Grant No. 62106127, National Science Foundation for Distinguished Young Scholars of China with No. 61825204 and Beijing Outstanding Young Scientist Program with No. BJJWZYJH01201910003011.

References

[1] Z. Zhang *et al.*, "Seccl: Securing collaborative learning systems via trusted bulletin boards," *IEEE Communications Magazine*, vol. 58, no. 1, pp. 47–53, 2020.

[2] A. D'Alconzo *et al.*, "A survey on big data for network traffic monitoring and analysis," *IEEE Transactions on Network and Service Management*, vol. 16, no. 3, pp. 800–813, 2019.

[3] R. Boutaba *et al.*, "A comprehensive survey on machine learning for networking: evolution, applications and research opportunities," *Journal of Internet Services and Applications*, vol. 9, no. 1, pp. 1–99, 2018.

[4] T. Zhang *et al.*, "Machine learning for end-to-end congestion control," *IEEE Communications Magazine*, vol. 58, no. 6, pp. 52–57, 2020.

[5] A. B. Arrieta *et al.*, "Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai," *Information Fusion*, vol. 58, pp. 82–115, 2020.

[6] M. Ridwan *et al.*, "Applications of machine learning in networking: A survey of current issues and future challenges," *IEEE Access*, 2021.

[7] Y. Zheng *et al.*, "Demystifying deep learning in networking," in *Proc. APNet*, 2018, pp. 1–7.

[8] Z. Meng *et al.*, "Interpreting deep learning-based networking systems," in *Proc. ACM SIGCOMM*, 2020, pp. 154–171.

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

[10] R. Inam *et al.*, "Explainable AI – how humans can trust AI," https://www.ericsson.com/en/reports-and-papers/white-papers/explainable-ai–how-humans-can-trust-ai, Accessed on 09.10.2021.

[11] C. Beliard *et al.*, "Opening the deep pandora box: Explainable traffic classification," in *Proc. INFOCOM*. IEEE, 2020, pp. 1292–1293.

[12] A. Morichetta *et al.*, "EXPLAIN-IT: towards explainable AI for unsupervised network traffic analysis," in *Proc. ACM Big-DAMA*, 2019, pp. 22–28.

[13] S. C. Sun *et al.*, "Approximate symbolic explanation for neural network enabled water-filling power allocation," in *Proc. VTC2020-Spring*. IEEE, 2020, pp. 1–4.

[14] W. Guo, "Partially explainable big data driven deep reinforcement learning for green 5g uav," in *Proc. ICC*. IEEE, 2020, pp. 1–7.

[15] A. Terra *et al.*, "Explainability methods for identifying root-cause of sla violation prediction in 5g network," in *Proc. GLOBECOM*. IEEE, 2020, pp. 1–7.

Tianzhu Zhang is a research engineer at Nokia Bell Labs. He received his B.S. degree from Huazhong University of Science and Technology, Wuhan, China, in 2012. Afterward, he received the M.S. degree in 2014 and the Ph.D. degree in 2017 from Politecnico di Torino, Turin, Italy. From 2017 to 2019, he was a PostDoc researcher at Telecom ParisTech and LINCS under a research grant from Cisco Systems. He joined Nokia Bell Labs in August 2020. His research interests include SDN, NFV, edge computing, and AI.

Han Qiu received the B.E. degree from the Beijing University of Posts and Telecommunications, Beijing, China, in 2011, the M.S. degree from Telecom-ParisTech (Institute Eurecom), Biot, France, in 2013, and the Ph.D. degree in computer science from the Department of Networks and Computer Science, Telecom-ParisTech, Paris, France, in 2017. He worked as a postdoc and a research engineer with Telecom Paris and LINCS Lab from 2017 to 2020. Currently, he is an assistant professor at Tsinghua University. His research interests include AI security, data security, and cloud computing.

Marco Mellia (M'97–SM'08-F'20) is a full professor at the Control and Computer Engineering Department of Politecnico di Torino. In 2002 he visited the Sprint Advanced Technology Laboratories working at the IP Monitoring Project (IPMON). In 2011 - 2013, he collaborated with Narus Inc, CA, working on traffic monitoring and cybersecurity system design. In 2015 and 2016, he collaborated with Cisco Systems to design cloud monitoring platforms based on machine learning. He is now the coordinator of the SmartData@PoliTO center on data science and machine learning, involving more than 50 colleagues and Ph.D. students. Prof. Mellia co-authored over 250 papers published in international journals and presented in leading conferences, all of them in communication networks.

Yuanjie Li is assistant professor at the Institute for Network Sciences and Cyberspace at Tsinghua University. He was a researcher at Hewlett Packard Labs from 2018 to 2020, and the co-founder of MobIQ Technologies from 2017 to 2018. He received his Ph.D. in Computer Science from UCLA in 2017, and B.E. in Electronic Engineering from Tsinghua University in 2012. His research interests are network systems and security, with a recent focus on mobile networking, intelligent wireless edge, and Internet-of-Things (IoT).

Hewu Li received his M.S. and Ph.D. degrees in computer science from Tsinghua University, in 2001 and 2004, respectively. He is currently an associate professor and assistant to the Dean of the Institute for Network Sciences and Cyberspace at Tsinghua University. He is also the director of the Wireless and Mobile Network Technology research laboratory, 2009 AsiaFI (Asia Future Internet) Wireless Chairman of the Mobile Working Group. His research interests include mobile and wireless networks mobile wireless network architecture, hybrid satellite-terrestrial networks, broadband wireless access technology, and mobility architecture in next-generation networks.

Ke Xu is currently full professor with the Department of Computer Science & Technology of Tsinghua University, Beijing, China. He has authored/co-authored more than 200 technical papers and holds 11 US patents in the research areas of next-generation Internet, blockchain systems, Internet of Things, and network security. He received the Ph.D. degree from the Department of Computer Science & Technology, Tsinghua University. He is a member of ACM and a Senior Member of IEEE. He is an Editor for the IEEE Internet of Things Journal. He is a Steering Committee Chair of IEEE/ACM International Symposium on Quality of Service.