Toward Native Explainable and Robust AI in 6G Networks: Current State, Challenges and Road Ahead

Claudio Fiandrino[a], Giulia Attanasio[a,b], Marco Fiore[a], Joerg Widmer[a]

[a] *IMDEA Networks Institute, Madrid, Spain*
[b] *Universidad Carlos III de Madrid, Spain*

Abstract

6G networks are expected to face the daunting task of providing support to a set of extremely diverse services, each more demanding than those of previous generation networks (e.g., holographic communications, unmanned mobility, etc.), while at the same time integrating non-terrestrial networks, incorporating new technologies, and supporting joint communication and sensing. The resulting network architecture, component interactions, and system dynamics are unprecedentedly complex, making human-only operation impossible, and thus calling for AI-based automation and configuration support. For this to happen, AI solutions need to be robust and interpretable, i.e., network engineers should trust the way AI operates and understand the logic behind its decisions. In this paper, we revise the current state of tools and methods that can make AI robust and explainable, shed light on challenges and open problems, and indicate potential future research directions.

Keywords: 6G networks, AI, explainable AI, robust AI.

---

1. Introduction

Fifth-generation (5G) networks are now entering a stable phase in terms of system architecture and commercial release, and the identification of the

Email addresses: claudio.fiandrino@imdea.org (Claudio Fiandrino), giulia.attanasio@imdea.org (Giulia Attanasio), marco.fiore@imdea.org (Marco Fiore), joerg.widmer@imdea.org (Joerg Widmer)

advanced features that will shape the evolution of 5G into the sixth generation (6G) of mobile network systems has already started [1]. Despite being in the early stages of conceptualization, some key aspects of the future infrastructure have been identified by the community: 6G will bring a paradigm shift from "connected things" to "connected intelligence," supporting even more stringent KPI requirements than 5G, and global coverage [2]. Therefore, there are strong expectations that Artificial Intelligence (AI) will permeate the 6G network infrastructure, allowing for much swifter and more effective decision-making in scheduling, control, and orchestration operations of the end-to-end communication systems [3]. Ultimately, this will allow 6G to support ambitious performance targets such as near-zero latency, apparent infinite capacity, and near 100% reliability and availability, to support new and diverse classes of innovative mobile services.

When applied to specific network functionalities, AI systems will either employ pre-trained models or adapt those at operation time. These models include machine and deep learning models, and are specialized in analyzing large data and identifying complex relations and patterns that extend beyond human knowledge. In a nutshell, this process happens by relating input data to outputs with stochastic processes. The deep learning models that are typically used to address wireless network problems are stochastic while traditional software is by nature deterministic. This implies that existing formal verification tools that are usually employed for testing software robustness are no longer valid [4]. To add another dimension to the challenge, deep learning models are regarded as black-box [5]. In other words, it is very hard to understand the underlying operation and the reasons why the models have taken certain actions [6].

In light of the considerations above, it is of paramount importance that AI becomes *trustworthy*, in the sense that AI models should be robust and explainable for humans to *trust* such non-deterministic systems. Given the growing interest in the matter, the landscape of regulations by national and international bodies is continuously evolving [7]. Among others, Article 13 of the EU Regulatory Framework for AI[1] states: "*AI systems should be sufficiently transparent, explainable, and well-documented.*" In human-in-the-loop scenarios, understanding how complex models operate is critical for system experts to perform root-cause analysis [8]. This applies to the vast majority of zero-touch network configuration and automation scenarios [9] under discussion within the ETSI ZSM (Zero-touch network and Service Management) group. Furthermore, to be trusted, AI models should be robust. Previous research has revealed that adding a small change to the inputs is sufficient to fool a classifier, e.g., the infamous tape strip over a speed limit sign that leads a classifier to accelerate and not to brake or, in the context of mobile networks, to misclassify wireless signals sent for authentication that are generated by non-legitimate users [10].

In this paper, we provide an up-to-date primer on robust and explainable AI for mobile networks. We outline and review existing tools, their applicability, and shortcomings to address 6G network challenges (§ 2). Next, we discuss how to enable robust and explainable AI in 6G networks and integrate it into the current network architecture models (§ 3). We then present a case study to expose the complexity of applying explainability concepts to a deep learning mobile traffic predictor based on real-world traffic data (§ 4). Finally, we draw conclusions and analyze future research directions (§ 5).

2. Tools and Methods for Explainable, Robust and Verifiable AI

This section presents background on explainable (§ 2.1) and robust AI (§ 2.2) and formal verification techniques (§ 2.3) for AI models. Next, it provides a discussion (§ 2.4) that highlights shortcomings of existing tools when applied to mobile networks.

2.1. Explainable AI

The growing interest in promoting trust in ICT systems has been addressed by regulatory bodies at different levels [7]. In the context of AI/ML, DARPA has

[1] Available online at: `https://bit.ly/3FATnNj` - Last accessed: 04/05/2022

introduced the Explainable AI (XAI) initiative to promote research around model interpretability to ultimately open up the AI/ML models' black-box behavior and make it more intelligible to humans [6]. Such initiative sparked the interest of the AI community and model interpretability is becoming an important feature as a basis for new designs. For example, Auric [11], a framework that is used by AT&T (a US mobile operator) to automatically configure base stations (known as eNB and gNB in LTE and 5G jargon respectively) parameters, is based on decision trees that offer good results in trading off accuracy and interpretability.

Despite some first results, XAI remains a wide-open research area. While some models like decision trees are easy to interpret and have already been utilized in practice [11], in the mobile network domain, the vast majority of AI/ML applications (e.g., routing, load balancing, and resource allocation) use much more complex AI models like deep learning models [5]. The computer vision and Natural Language Processing (NLP) domains received comparatively more attention than the domain of time-series analysis because of the rich semantics of the inputs that are intuitive to humans. By setting to zero a given set of pixels of an image (perturbation) [6], it is possible to visually understand their contribution to a model (e.g., which lung regions from X-RAY images are important to detect COVID-19 [12]). Using such an approach for time series is technically feasible at the cost of disrupting the temporal dependencies. Explainability allows to better comprehend how models operate, thereby allowing to strengthen robustness and resiliency [13]. At the same time, assessing robustness and resilience with specific perturbations allows to understand better which input patterns are prone to weaken the model accuracy. Layer-wise Relevance Propagation (LRP), DeepLIFT, Local Interpretable Model-Agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP) are existing methods for interpretability [14]. Unlike LRP, all the other methods resort to perturbing the inputs to measure the accuracy drop with respect to the original model. By contrast, LRP uses the neural network weights and the activations that have been created with the forward-pass to propagate back the output until the input layer. For this reason, LRP can not be applied to any model out of the box like SHAP: there are existing implementations for popular models like LSTM, and bi-directional LSTM. TSVis [15], Long-Short Term Memories (LSTM)-Vis [16] and Sequence to Sequence (Seq2Seq)-Vis [17] are visualization tools that apply respectively to CNN, LSTM and Seq2Seq learning models and aim at tracking the hidden state changes. The latter two are conceived as a tool for NLP applications.

2.2. Adversarial Machine Learning

Perturbation is key to test robustness and resilience against adversarial attacks. Adversarial Machine Learning (AML) comprises several techniques that build on this concept and ultimately define the trustworthiness of an AI model. Seminal works like [18] revealed that adding a small change to the inputs is sufficient to fool a classifier. Attacks performed against an AI model can be white-box, gray-box, or black-box, depending on the amount of information the attacker has about the model itself. The first category assumes that the adversary has full knowledge of the training data, model architecture, and parameters, the latter none, and gray-box attacks assume partial knowledge. Known attacks on time-series are modifications of attacks originally designed for images like the Fast Gradient Sign Method (FGSM) and its iterative version Basic Iterative Method (BIM) [19]. Both generate a crafted input that resembles the original one, but the values of its elements are equal to the sign of the elements of the gradient of the cost function. This is enough to increase the classification/forecast error.

2.3. Formal Verification for AI

Besides being able to counteract adversaries, to be fully trustworthy AI would require formal verification. In early 2000, formal verification boosted software development with systematic bug detection in code, vulnerability analysis, threat analysis, and run-time monitoring. However, applying the same concepts to AI is challenging because i) many AI models like deep learning ones are stochastic by nature as opposed to the deterministic nature of computing systems, and ii) the role of data becomes crucial [4]. A learning model is trained on a particular dataset and it is well known that adding additional input data usually leads to an increase of accuracy. Applying formal methods to AI is not new and a recent article surveys the tools and methods proposed so far [20]. Those applicable to neural networks are categorized into complete and incomplete formal methods. The former methods suffer from scalability but are sound, i.e., they can report if a given specification holds or not. In contrast, the latter scales better at the cost of reporting false positives. Complete methods can be further categorized into satisfiability modulo theory (SMT) and mixed-integer linear programming (MILP) based methods. SMT boils down the verification problem to a constraint satisfiability problem: if the modeled constraints can be satisfied, then the property is not verified. MILP-based methods transform the verification problem into a MILP-one. If the objective function can be maximized or minimized, then the property is not verified because it exists at least one counter-example.

2.4. Discussion

The existing systems for explainability and trustworthiness outlined above have several shortcomings when applied to the mobile network domain, as follows.

First, while the interpretability and visualization tools are relevant, they fail to explain at a deeper level the model operation. Just highlighting that a given load pattern at a given time triggers the activation of many neurons does not explain how important this is in relation to the nature of input data that produced such behavior. Visualization tools should be extended and coupled with data mining techniques like Gramian Angular Field or Markov Transition Field [21] to fully comprehend and exploit the nature of the input data. In addition to understanding the reason for producing a given output, a comprehensive tool should also unveil which patterns are responsible for the errors.

Second, the existing time-series attacks do not consider the specific requirements of mobile network inputs like traffic load or channel propagation. For instance, the traffic load cannot be a negative value for a given base station. Or, jamming multiple transmissions at a base station to decrease the observed load is an extremely hard task in practice, which requires knowledge of the exact timing of data transmission and of the time-varying characteristics of the channels from the base stations to the users and from the jammers to the users. XAI and AML tools can be used jointly, for example to understand if a model over-learns outliers and if these outliers are legitimate or rather forged ad-hoc.

Third, the existing AI formal verification tools are conceived for simple neural networks and model parameters, and have all been tested only for popular image datasets (e.g., MNIST, CIFAR-10). Therefore, they require a significant extension to accommodate complex architectures that are commonly applied to address mobile networking problems [22].

3. Explainability and Robustness: Integration in Next Generation Mobile Networks

We now discuss how techniques for AI robustness, explainability, and verification, based on and possibly extending the approaches presented in the previous Section, can be integrated in the 6G network architecture.

While the 6G Radio Access Network (RAN) will significantly evolve from 5G, at least in its earlier deployment, the 6G core will retain part of the concepts and functionalities of the 5G core network. The reason is twofold. First, the 5G core network is radically different from previous generations' core networks and Mobile Network Operators (MNO) are likely not willing to make a significant capital expenditure for a new change. Second, the 5G core network was designed to adhere to cloud-native and service-based architecture principles, which make it easy to extend it to support new functionalities like location-based analytics [23].

Fig. 1 shows the components of the mobile network architecture with 3GPP standardized functions in the user-plane, control-plane, and radio part [24].[2] For 5G, initial support for AI is provided by the Network Data Analytics Function

[2] A comparison with previous generation architectures, and a thorough presentation of the purpose of each of the 5G standardized functions is out the scope of this work. We refer the interested reader to the complete description in [24].

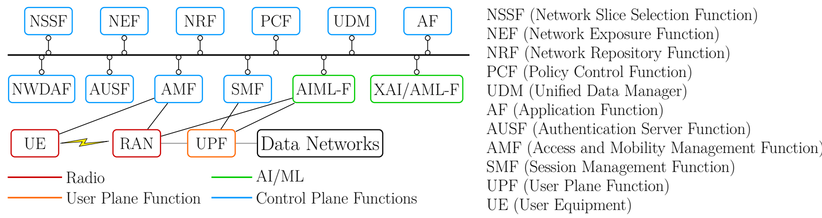

Figure 1: Integration of AI/ML with tools to support robustness and explainability in the 6G network architecture

(NWDAF) in the core and by the Radio Network Information Base (RNIB) in the RAN [25]. Beyond 5G and 6G network architectures will likely comprise functions to exploit AI/ML as a service to optimize specific mechanisms and functionalities. The recently proposed AI/ML Platform (AIMLP) is an example on how to implement AI/ML as a service [26]. Specifically, the AI/ML-Function (AIML-F) in Fig. 1 (that can be mapped to AIMLP) will contain pre-trained learning models ready to be used by other functions (once trained, learning models can be exported with information on all the weights in *hd5* format). Similarly to the AIML-F, a new function will host ready-to-apply tools for assessing the robustness of the AI models and provide explanations of their execution (the XAI/AML-F in Fig. 1). Standard interfaces will provide AI models and human-in-the-loop capabilities to access and execute XAI/AML tools on the data or model of interest. For this to happen, the computing platform of an MNO needs to adapt to accommodate computing- and memory-intensive AI tasks. While the community has well highlighted this need for executing all the operations involved with an ML pipeline (e.g., collection of measurement data, distributed/centralized model training and inference execution, actuation with a change of network mechanism policy), also XAI/AML tools are computationally expensive. The case study we present in the next Section provides a practical example of the substantial computational requirements of XAI methods for networking.

4. Case Study

We consider a specific case study, which we employ to set forward the path to address one of the shortcomings highlighted in Section 2.4, i.e., that of ensuring deeper model explainability with the help of time series mining techniques. For this, we focus on mobile traffic forecasting, one of the most popular applications of AI for mobile networking. Next, we present the dataset used (§ 4.1) and discuss how to ensure deeper model explainability and the associated execution time, CPU and memory footprint (§ 4.2).

4.1. The Dataset

For our experiments, we rely on a measurement dataset of real-world traffic collected in a production 4G network serving a major metropolitan region in Europe. The data consists of information on the traffic volume generated by a set of target mobile applications, including popular services like Apple iCloud, Facebook, Netflix, and Whatsapp, among others, at each eNB. The traffic maps to the demand of the whole user base of the operator in the region, which has a market share of more than 30% there.

The data was collected via commercial passive probes that tap into interfaces of the Gateway GPRS Support Nodes (GGSNs) and the Packet Data Network Gateways (PGWs), monitor individual flows, and perform traffic classification using Deep Packet Inspection (DPI) and proprietary fingerprinting solutions. The processing of flow-level captures into per-minute traffic volumes at each eNB occurred in the secure premises of the network operator, under the supervision of the local Data Protection Officer (DPO). We only had access to the depersonalized aggregates for our study, in compliance with applicable international regulations.

Overall, the dataset comprises the per-eNB time series of 23 mobile services at the granularity of three minutes. All the time series cover the same period of 11 weeks in the fall of 2019.

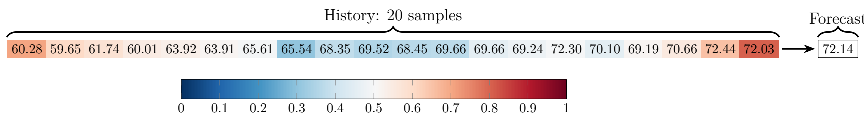

Figure 2: Model explanation with LRP

4.2. Explaining DL Models for Traffic and Capacity Forecasting

The univariate time series of the service-level demand aggregated over all eNBs is fed to a deep neural network. The forecasting task then maps to anticipating the future load in the region of interest. To this end, we employ an LSTM layer with 200 memory cells followed by a fully connected output layer with a single hidden unit for the actual prediction[3]. The deep neural network receives a history of past observations $T_n \in \mathcal{T} = t_{n-k+1}, t_{n-k+2}, \ldots, t_n$ of the input feature, *i.e.*, load expressed in MB/min, and aims at forecasting the load at the time instant t_{n+1}. For our analysis, we set $k = 20$ (*i.e.*, 1 hour of the time series as each sample characterizes the load over 3 minutes) and $k = 120$ (*i.e.*, 6 hours). We train our network over 9 weeks and we test over the last two weeks of the dataset. The model is trained using MAE as the loss function and the Adam optimizer with a learning rate of 0.0001 during 470 epochs. We verify that the model outperforms by 15% in terms of Mean Absolute Error a naive predictor whose forecast at t_{n+1} corresponds to the load at t_n. We perform our explainability analysis on the test set by using both LRP and SHAP methods. For LRP we use the implementation by Warnecke et al. [27] with $\epsilon = 10^{-3}$. Instead, for SHAP we use the open-source implementation by Lundberg et al. [28] including the DeepExplainer method.

We now compare the model operation explained by LRP and SHAP. Our methodology is as follows: we explain how the model predicts the value t_{n+1} of the load time-series $\mathcal{T} = t_1, t_2, \ldots, t_n$, using as history the last $k = 20$ values $T_n \subset \mathcal{T} = t_{n-k+1}, t_{n-k+2}, \ldots, t_n$. Both methods identify a general trend in the way in which the model works: recent and old samples contribute positively to the forecast, while samples in the center of T_n are less relevant. Fig. 2 and Fig. 3

[3]The neural network architecture was selected on the basis of extensive tests.

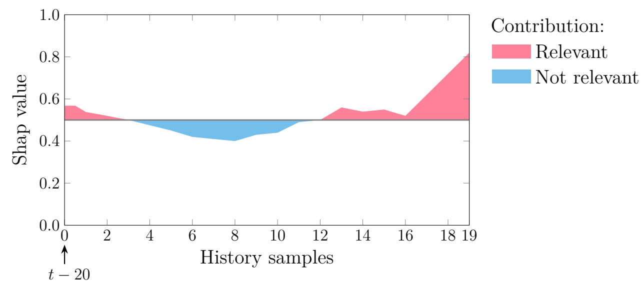

Figure 3: Model explanation with SHAP

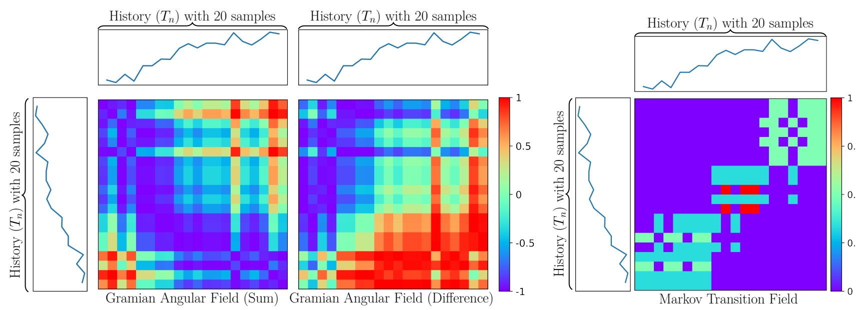

Figure 4: Gramian Angular Fields and Markov Transition Field applied to our data

show respectively an example of LRP and SHAP explanations.

After having identified that the both XAI methods provide a similar explanation regarding how the LSTM model operates, we mine the input data with two techniques that encode time-series as images: the Gramian Angular Field (GAF) in its sum and difference forms and the Markov Transition Fields (MTS) [21]. Fig 4 shows an example of application of GAF and MTS over a generic $T_n \subset \mathcal{T}$. While the MTS does not seem to provide any deeper explanation, the GAF does shed some light on the model operation. We know that samples at the extremities of T_n are highly relevant to predict t_{n+1}. By focusing on the GAF (sum, the most left plot), we can appreciate that at the extremities of T_n (bottom left and top right part of the first plot in Fig 4), many values are positive and many peak at 1. This means for the model, the most relevant samples of T_n are those values of load that are either very low (these are also old values) or very high (the most recent values of T_n).

We now characterize how computing intensive it is to execute the XAI tools. Table 1 and Table 2 show execution time and resource utilization for

LRP and SHAP, respectively, for machines with different hardware capabilities. Specifically, we test an Intel(R) Core(TM) i7-6800K CPU @ 3.40GHz (12 cores), equipped with 64 GB of RAM (Server 1), an AMD Ryzen 9 5950X Processor (16 cores) with 64 GB of RAM (Server 2), an 11th Gen Intel(R) Core(TM) i9-11900K @ 3.50GHz (16 cores) with 64 GB of RAM (Server 3) and an Intel(R) Xeon(R) Gold 6240R CPU @ 2.40GHz (97 cores) equipped with 264 GB of RAM (Server 4). Values of mean and standard deviation for CPU and memory consumption are obtained with the `glances` tool, with 1 sample every 3 seconds throughout the execution (shown in the second and third columns of Table 2). This analysis shows that for an increasing number of past observations fed to the explainability method, both the CPU usage and the allocated memory increase too. We registered an increment in the CPU usage of 20-100% while the allocated memory almost doubles. In particular, the measured memory allocation is more stable compared to CPU usage. The reason may be related to the fact that memory is preallocated by the explainability method and released to the Operating System (OS) at the end of the process execution, while the CPU is allocated and reclaimed by the OS when is needed.

Server	Time		Cpu				Memory			
	1 h	6 h	1 h		6 h		1 h		6 h	
			Mean	Std	Mean	Std	Mean	Std	Mean	Std
Server 1	4.39 s	15.43 s	31.7 %	11.2 %	58.2%	39.8%	2.1 %	0.1 %	4.5%	1.5 %
Server 2	1.44 s	6.04 s	40.3 %	35.5 %	88.4 %	46.2 %	5.2 %	0.1 %	5.6 %	0.2 %
Server 3	1.98 s	8.50 s	35.9 %	31.2 %	43.9 %	42.5 %	4.4 %	0.1 %	5.3 %	0.3 %
Server 4	2.07 s	11.26 s	55.0 %	17.3 %	83.3 %	21.9 %	1.8 %	0.1 %	2.4 %	0.2 %

Table 1: Profiling resource utilization of LRP on different machines

5. Concluding Remarks

In this paper, we looked at how to promote trustworthiness in the way future AI systems will be applied to next-generation mobile networks. For this

Server	Time		Cpu				Memory			
	1 h	6 h	1 h		6 h		1 h		6 h	
			Mean	Std	Mean	Std	Mean	Std	Mean	Std
Server 1	11.74 s	62.00 s	29.8 %	8.3 %	36.8 %	16.7 %	1.5 %	0.1 %	2.6 %	0.2 %
Server 2	6.22 s	33.30 s	30.9 %	24.6 %	43.8 %	11.4 %	1.3 %	0.1 %	2.4 %	0.1 %
Server 3	6.07 s	31.65 s	13.5 %	11.3 %	25.3 %	4.7 %	1.4 %	0.2 %	2.3 %	0.1 %
Server 4	10.52 s	58.28 s	7.66 %	3.82 %	12.3 %	3.8 %	1.6 %	0.2 %	2.4%	0.1 %

Table 2: Profiling resource utilization of SHAP on different machines

to happen, AI should become *explainable*, *robust* and *verifiable*. After having outlined the existing landscape of tools that enable the aforementioned properties, we discussed the shortcomings if applied directly to mobile networking problems.

For the case study of mobile traffic prediction, we showed the benefit of superior interpretability if the explanations that the XAI tools provide on the model operation are coupled with techniques that mine input data. Further work should be done in this area: understanding which mining techniques offer suitable explanations in more complex cases than the one presented in this work is nontrivial. Correlation or more complex mathematical tools like causal analysis could help in this regard to connect the dots. Besides understanding the reasons for producing a given output, a comprehensive tool should also unveil which patterns are responsible for the errors.

Acknowledgment

This work is partially supported by a Juan de la Cierva grant from the Spanish Ministry of Science and Innovation (IJC2019-039885-I), by the European Union's Horizon 2020 research and innovation programme under grant agreement no.101017109 "DAEMON", by the Atracción de Talento Investigador grant number 2019-T1/TIC-16037 NetSense funded by the Comunidad de Madrid, and by the Madrid Regional Government through the TAPIR-CM program (S2018/TCS-4496).

References

[1] C. D. Alwis, A. Kalla, Q.-V. Pham, P. Kumar, K. Dev, W.-J. Hwang, M. Liyanage, Survey on 6G frontiers: Trends, applications, requirements, technologies and future research, IEEE Open Journal of the Communications Society 2 (2021) 836–886.

[2] I. F. Akyildiz, A. Kak, S. Nie, 6G and beyond: The future of wireless communications systems, IEEE Access 8 (2020) 133995–134030.

[3] S. Muhammad K., R. Luca, B. M. Majid, K. Istvan Z., A. Mohamad, G. Mohsen, Artificial intelligence for 6G networks: Technology advancement and standardization, arXiv:2204.00914 (2022).

[4] J. M. Wing, Trustworthy AI, Commun. ACM 64 (10) (2021) 64–71.

[5] C. Zhang, P. Patras, H. Haddadi, Deep learning in mobile and wireless networking: A survey, IEEE Communications Surveys Tutorials 21 (3) (2019) 2224–2287.

[6] W. Samek, T. Wiegand, K.-R. Müller, Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models, arXiv:1708.08296 (2017).

[7] S. E. Middleton, E. Letouzé, A. Hossaini, A. Chapman, Trust, regulation, and human-in-the-loop AI: Within the european region, Commun. ACM 65 (4) (2022) 64–68.

[8] A. Terra, R. Inam, S. Baskaran, P. Batista, I. Burdick, E. Fersman, Explainability methods for identifying root-cause of SLA violation prediction in 5G network, in: Proc. of IEEE GLOBECOM, 2020, pp. 1–7.

[9] B. Angui, R. Corbel, V. Q. Rodriguez, E. Stephan, Towards 6G zero touch networks: The case of automated cloud-RAN deployments, in: Proc of IEEE CCNC, 2022, pp. 1–6.

[10] Y. Shi, K. Davaslioglu, Y. E. Sagduyu, Generative adversarial network in the air: Deep adversarial learning for wireless signal spoofing, IEEE Transactions on Cognitive Communications and Networking 7 (1) (2021) 294–303.

[11] A. Mahimkar, A. Sivakumar, Z. Ge, S. Pathak, K. Biswas, Auric: Using data-driven recommendation to automatically generate cellular configuration, in: Proc. of the ACM SIGCOMM, 2021, p. 807–820.

[12] A. Malhotra, S. Mittal, P. Majumdar, et al., Multi-task driven explainable diagnosis of COVID-19 using chest X-ray images, Pattern Recognition 122 (2022) 108243.

[13] T. Rojat, R. Puget, D. Filliat, J. Del Ser, R. Gelin, N. Díaz-Rodríguez, Explainable artificial intelligence (XAI) on timeseries data: A survey, arXiv:2104.00950 (2021).

[14] U. Schlegel, H. Arnout, M. El-Assady, D. Oelke, D. A. Keim, Towards a rigorous evaluation of XAI methods on time series, in: Proc. of IEEE ICCVW, 2019, pp. 4197–4201.

[15] S. A. Siddiqui, D. Mercier, M. Munir, A. Dengel, S. Ahmed, TSViz: Demystification of deep learning models for time-series analysis, IEEE Access 7 (2019) 67027–67040.

[16] H. Strobelt, S. Gehrmann, H. Pfister, A. M. Rush, LSTMVis: A tool for visual analysis of hidden state dynamics in recurrent neural networks, IEEE Transactions on Visualization and Computer Graphics 24 (1) (2018) 667–676.

[17] H. Strobelt, S. Gehrmann, M. Behrisch, A. Perer, H. Pfister, A. M. Rush, Seq2seq-vis: A visual debugging tool for sequence-to-sequence models, IEEE Transactions on Visualization and Computer Graphics 25 (1) (2019) 353–363.

[18] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, R. Fergus, Intriguing properties of neural networks, in: International Conference on Learning Representations (ICLR), 2014.

[19] G. R. Mode, K. A. Hoque, Adversarial examples in deep learning for multivariate time series regression, in: Proc. of IEEE AIPR, 2020, pp. 1–10.

[20] C. Urban, A. Miné, A review of formal methods applied to machine learning, arXiv:2104.02466 (2021).

[21] Z. Wang, T. Oates, Spatially encoding temporal correlations to classify temporal data using convolutional neural networks, arXiv:1509.07481 (2015).

[22] D. Bega, M. Gramaglia, M. Fiore, A. Banchs, X. Costa-Pérez, DeepCog: Optimizing resource provisioning in network slicing with AI-based capacity forecasting, IEEE JSAC 38 (2) (2020) 361–376.

[23] S. Bartoletti, L. Chiaraviglio, S. Fortes, et al., Location-based analytics in 5G and beyond, IEEE Communications Magazine 59 (7) (2021) 38–43.

[24] 5GPPP Architecture Working Group and others, View on 5G architecture: Version 4.0 (2021).

[25] D. Bega, M. Gramaglia, R. Perez, M. Fiore, A. Banchs, X. Costa-Pérez, AI-based autonomous control, management, and orchestration in 5G: From standards to algorithms, IEEE Network 34 (6) (2020) 14–20.

[26] C. Papagianni, J. Mangues-Bafalluy, P. Bermudez, et al., 5Growth: AI-driven 5G for automation in vertical industries, in: Proc. of IEEE EuCNC, 2020, pp. 17–22.

[27] A. Warnecke, D. Arp, C. Wressnegger, K. Rieck, Evaluating explanation methods for deep learning in security, in: Proc. of IEEE EuroS&P, 2020, pp. 158–174.

[28] S. M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, in: Advances in Neural Information Processing Systems, 2017, pp. 4765–4774.

Biographies

Claudio Fiandrino is a postdoctoral researcher at IMDEA Networks Institute. He obtained his Ph.D. degree from the University of Luxembourg in 2016. His research focuses on mobile networks in the areas of AI-driven network optimization and edge computing. Claudio was awarded with two Spanish Juan de la Cierva grants and 3 Best Paper Awards for his research.

Giulia Attanasio is a Ph.D. student in Telematics at the University of Carlos III. Her research interest lies between machine learning and mobile networks communications. Previously, she completed an MSc in Communications and Computer Networks Engineering and a BS in Telecommunications Engineering at the Polytechnic University of Turin.

Marco Fiore is a Research Associate Professor at IMDEA Networks Institute, Spain, where he leads the Networks Data Science group on research activities at the interface of mobile networking and applied data science. He previously held tenured positions in France and Italy, and was a Marie Curie fellow and a Royal Society visiting research fellow.

Joerg Widmer is Research Professor and Research Director of IMDEA Networks in Madrid, Spain. His research focuses on wireless networks, in particular millimeter-wave communications. Joerg Widmer is an IEEE Fellow and Distinguished Member of the ACM, and was awarded an ERC consolidator grant, the Friedrich Wilhelm Bessel Research Award of the Alexander von Humboldt Foundation, a Mercator Fellowship of the German Research Foundation, and a Spanish Ramon y Cajal grant.