Research Trends on XAI for Open RAN

Seungyeop Song, Heejae Park, Seongryool Wee, and Laihyuk Park
Department of Computer Science and Engineering, Seoul National University of Science and Technology, Seoul, 01811, Korea
Email: {sysong, prkhj98, holylaw, lhpark}@seoultech.ac.kr

***Abstract*—RAN Intelligent Controller (RIC) in O-RAN leverages Artificial Intelligence (AI) to enhance service quality in dynamic network environments. However, AI models have a black-box nature, making their internal operations difficult to interpret. Therefore, the adoption of AI models in RIC requires techniques that provide a clear understanding of model behavior and decision-making processes. eXplainable AI (XAI) has emerged as a core technology that can solve this problem. This paper examines the O-RAN architecture and recent research on the integration of XAI techniques into O-RAN systems.**

***Index Terms*—O-RAN, XAI, research trends**

I. Introduction

RAN Intelligent Controller (RIC) in O-RAN leverages Artificial Intelligence (AI) to enhance service quality in dynamic network environments [1]. However, AI models have a black-box nature, making their internal operations difficult to interpret. Therefore, the adoption of AI models in RIC requires techniques that provide a clear understanding of model behavior and decision-making processes.

In this context, eXplainable AI (XAI) has emerged as a key technology that provides the transparency and reliability necessary for network operation. It achieves this by interpreting the internal operating principles of AI models and identifying the contribution of input features [2].

Recent studies focus on the integration of XAI into O-RAN for improved decision-making and system efficiency. This paper reviews the O-RAN architecture and recent research on the integration of XAI techniques into O-RAN systems.

II. O-RAN Architecture

Fig. 1 shows the O-RAN Architecture [3]. The service management and orchestration (SMO) performs all orchestration procedures for monitoring and controlling O-RAN components. The RIC performs network management and control and is divided into two types based on latency requirements. Near-RT-RIC controls the RAN in real-time and supports the execution of xApp. Non-RT-RIC operates in non-real-time and supports the execution of rApp.

In the O-RAN architecture, the traditional base station is disaggregated into an open centralized unit (O-CU), an open distributed unit (O-DU), and an open radio unit (O-RU). In addition, the O-CU is divided into a control plane (CP) and a user plane (UP). Each component is interconnected through standardized interfaces. The A1 interface connects the Non-RT RIC and the Near-RT RIC. The E2 interface connects the Near-RT RIC to the E2 node such as the DU and the CU. The O1 interface connects the SMO and the O-RAN components,

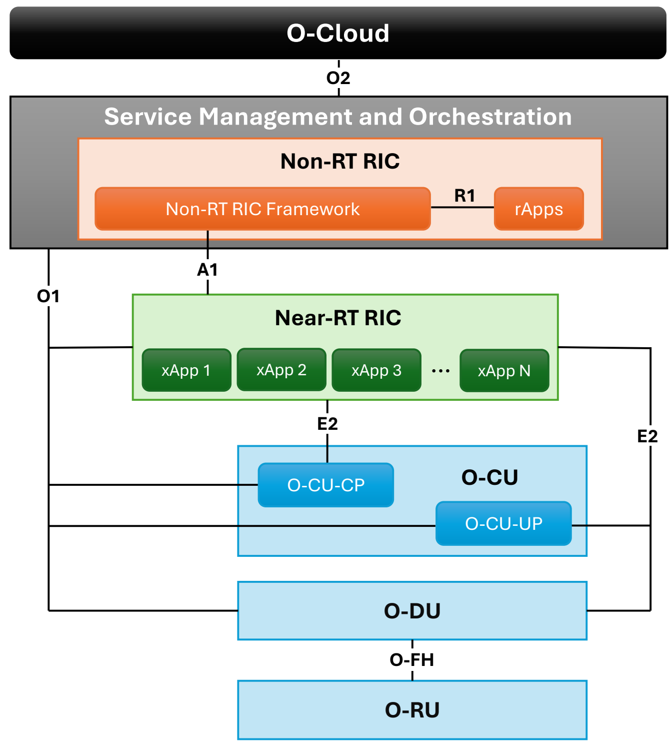

Fig. 1. O-RAN architecture

and the O2 interface connects the SMO and the O-Cloud. The O-FH (O-RAN FrontHaul) interface connects the O-DU and the O-RU, and the R1 interface connects the Non-RT RIC framework and the rApp.

III. Research Trends

The authors in [1] proposed EXPLORA (AI/ML EXPLainability for the Open RAN) that integrates XAI into the RIC. The technology offers an explainable representation of the decision-making process in DRL-based xApps and rApps. Utilizing an attributed graph-based technique, it visually elucidates the relationship between the DRL agent's actions and the corresponding input state. It was prototyped and evaluated on Colosseum, a large-scale wireless network emulator. Experimental results show that DRL control using EXPLORA achieved up to 4% improvement in median transmission bitrate and 10% gain at the tail.

In [4], the authors proposed an Explanation-Guided Federated Learning (EGFL) framework based on XAI for reliable control of 6G networks. The model used datasets generated by an O-RAN simulator, which emulates realistic network conditions, including diverse traffic characteristics and user mobility patterns. In the training process, the model ensures trustworthy predictions by exploiting XAI strategies via Jensen-Shannon (JS) divergence. The comprehensiveness score is used to measure and validate the faithfulness of the explanations quantitatively. Compared to existing Kullback-Leibler (KL) divergence-based methods or unconstrained federated learning methods, EGFL achieved relative improvements of more than 50% and 25%, respectively, in explanatory faithfulness and fairness.

The authors in [5] aim to overcome the limitations of fixed protocols in 6G O-RAN network slicing environments. They propose an explainable communication framework based on multi-agent deep reinforcement learning, called the Standalone Explainable Protocol (STEP). STEP integrates information bottleneck (IB) theory with deep Q-network (DQN) learning concepts. It also implements a stochastic bottleneck layer inspired by variational autoencoders (VAEs) to extract concise and semantically meaningful communication messages. Experimental results show that STEP outperforms the predefined protocol method and conventional multi-agent deep reinforcement learning (MADRL) approaches. It achieves up to a 6.06-fold reduction in inter-slice conflicts, a 1.4-fold improvement in resource utilization, and a 3.5-fold decrease in latency.

To address control traffic overhead on the E2 interface caused by KPI requests in O-RAN environments, the authors in [6] proposed a method that selects only important KPIs using SHAP-based XAI techniques. SHAP quantifies the extent to which each KPI contributes to the predictions of the machine learning model. This enables retraining of the model using only the KPIs that have a substantial impact on its performance. Experimental results show that using only the top contributing KPIs instead of the full set results in an accuracy loss of less than 7%, while reducing control traffic overhead by up to 33%.

IV. Conclusion

In this paper, the O-RAN architecture and research trends on XAI are reviewed. Recent studies demonstrate that the application of XAI techniques enhances control efficiency, resource utilization, and predictive performance within O-RAN environments. Future research will investigate the application of XAI to resource allocation in O-RAN-based V2X systems.

Acknowledgment

This work was partly supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (RS-2025-16070295) and IITP (Institute for Information & Communications Technology Planning & Evaluation) grant funded by the Korea government(MSIT) (No.RS-2024-00437252, Development of anti-sniffing technology in mobile communication and AirGap environments).

References

[1] C. Fiandrino, L. Bonati, S. D'Oro, M. Polese, T. Melodia, and J. Widmer, "Explora: Ai/ml explainability for the open ran," *Proceedings of the ACM on Networking*, vol. 1, no. CoNEXT3, pp. 1–26, 2023.

[2] B. Brik, H. Chergui, L. Zanzi, F. Devoti, A. Ksentini, M. S. Siddiqui, X. Costa-Pèrez, and C. Verikoukis, "Explainable ai in 6g o-ran: A tutorial and survey on architecture, use cases, challenges, and future research," *IEEE Communications Surveys & Tutorials*, 2024.

[3] M. Polese, L. Bonati, S. D'oro, S. Basagni, and T. Melodia, "Understanding o-ran: Architecture, interfaces, algorithms, security, and research challenges," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 2, pp. 1376–1411, 2023.

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

[5] F. Rezazadeh, H. Chergui, S. Siddiqui, J. Mangues, H. Song, W. Saad, and M. Bennis, "Intelligible protocol learning for resource allocation in 6g o-ran slicing," *IEEE Wireless Communications*, vol. 31, no. 5, pp. 192–199, 2024.

[6] C. Tassie, B. Kim, J. Groen, M. Belgiovine, and K. R. Chowdhury, "Leveraging explainable ai for reducing queries of performance indicators in open ran," in *ICC 2024-IEEE International Conference on Communications*. IEEE, 2024, pp. 5413–5418.