AI-RAN 기술 동향

양현종, 김유민, 최재현, 이현지*
서울대학교, *포항공과대학교

{hjyang, yumin0107, jhchoi0226}@snu.ac.kr , *leehji@postech.ac.kr

Trends in AI-RAN

Hyun Jong Yang, Yumin Kim, Jaehyun Choi, Hyunji Lee *
Seoul National Univ., *POSTECH

Abstract

This survey provides a concise overview of recent advancements in artificial intelligence applied to radio access networks (AI-RAN), emphasizing the emergence of wireless foundation models. It highlights key industry players, alliance activities, and cutting-edge research that collectively drive the evolution toward intelligent, AI-native wireless networks. The paper also discusses foundational model architectures and their potential to enhance network adaptability, efficiency, and automation in next-generation wireless systems.

Ⅰ. Introduction

The convergence of artificial intelligence (AI) and wireless communications is rapidly transforming the design and operation of next-generation radio access networks (RAN). As the demand for flexible, efficient, and autonomous wireless systems grows, AI-driven approaches—particularly those leveraging large-scale foundation models—are emerging as key enablers for intelligent network management, resource optimization, and real-time adaptation. Recent advances in wireless foundation models, inspired by breakthroughs in natural language processing and computer vision, promise to deliver universal, data-driven representations that can generalize across diverse wireless environments and tasks. This survey provides a concise overview of the latest developments in AI-RAN, with a special focus on the rise of wireless foundation models, their architectures, applications, and potential to shape the future of wireless networks.

II. What is the AI-RAN Alliance?

The AI-RAN Alliance [1], launched at Mobile World Congress 2024, is a global consortium of leading technology companies, telecom operators, and academic institutions dedicated to accelerating the integration of artificial intelligence into radio access networks (RAN). Its mission is to drive innovation, enhance network efficiency, and lay the foundation for AI-native 6G networks, emphasizing real-world validation and collaborative development over formal standardization. The Alliance is structured around three main working groups—AI-for-RAN, AI-and-RAN, and AI-on-RAN—and has rapidly grown to include 75 members from 17 countries. Key achievements include high-profile technology demonstrations at MWC, the launch of endorsed labs with industry and academic partners, and the introduction of flagship initiatives in data management and test methodologies. The recent activities of AI-RAN Alliance are summarized in Table I.

Ⅲ. Main Players of AI-RAN

The AI-RAN Alliance brings together leading global players from telecom, technology, and academia to accelerate the integration of AI into radio access networks. Founding members include NVIDIA, SoftBank, Samsung Electronics, Ericsson, Nokia, Microsoft, T-Mobile US, and Arm, with additional participation from

AWS, DeepSig, and major operators like KT, SK Telecom, and Globe Telecom. NVIDIA drives the Alliance's technical leadership, providing AI hardware and platforms for real-time inferencing and RAN automation. SoftBank, with its Research Institute of Advanced Technology, is central to the Alliance's strategy, with Dr. Alex Jinsung Choi serving as chair and advocating for AI-native 6G networks. Samsung and Ericsson contribute to AI-driven RAN orchestration and energy-efficient network solutions. Nokia and Microsoft focus on AI-based optimization and cloud-native RAN architectures. T-Mobile US and other operators are key to real-world deployment and validation of AI-RAN technologies. DeepSig and academic partners like Northeastern University support research and proof-of-concept development.

IV. Recent Papers on AI-RAN

Recent publications on AI-RAN highlight rapid progress in integrating advanced AI—such as deep learning, reinforcement learning, and large language models—into radio access networks to enable real-time, adaptive optimization for 6G and beyond. Key research themes include learning-based resource scheduling, traffic prediction, anomaly detection, and network slicing, with emerging paradigms like multi-agent systems, digital twins, and LLM-driven autonomous control moving RANs toward greater autonomy and cognition [2]. Studies emphasize the need for generalizable, interpretable AI models robust to dynamic wireless environments, as well as distributed architectures for edge coordination and efficient real-time inference. Industry reports and field trials show that current AI-RAN deployments primarily target operational efficiency and cost reduction, with operators like China Mobile and Verizon reporting significant improvements in mean time to repair, energy savings, and throughput using AI-powered solutions [3]. Major vendors and operators are also exploring AI-driven orchestration, edge-cloud collaboration, and the integration of AI-native functions into O-RAN architectures, while new entrants and established players alike are actively developing testbeds, open datasets, and benchmarking platforms to accelerate real-world adoption [4].

V. Wireless Foundation Model

Recent research on wireless foundation models focuses on developing large, generalizable AI models—often transformer-based—that can learn rich, universal representations from diverse wireless data, enabling improved performance across a variety of downstream wireless communication and sensing tasks. The "Large Wireless Model (LWM)" is the first such foundation model for wireless channels, pre-trained in a self-supervised manner and shown to enhance performance in complex and data-limited scenarios by generating contextualized channel embeddings [5]. Another work introduces WiMAE (Wireless Masked Autoencoder) and ContraWiMAE, multi-task foundation models that combine masked autoencoding and contrastive learning to capture both structural and discriminative channel features, outperforming existing baselines in adaptability and data efficiency [6]. The topic is gaining momentum, as reflected in upcoming dedicated workshops at major conferences like IEEE GLOBECOM 2025 [7].

ACKNOWLEDGMENT

This work was partially supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2023-00250191), and partially by the IITP(Institute of Information & Coummunications Technology Planning & Evaluation)-ITRC(Information Technology Research Center) grant funded by the Korea government(Ministry of Science and ICT)(IITP-2025-2021-0-02048)

참 고 문 헌

[1] AI-RAN Alliance, "AI-RAN Alliance," 2025. [Online]. Available: https://ai-ran.org. Accessed: May 20, 2025.

[2] *Journal of Information and Intelligence*, “Special Issue on Artificial Intelligence-Native Radio Access Networks (AI-RAN): Foundations, Methodologies and Applications,” May 16, 2025.

[3] S. Pongratz, “MWC 2025: RAN, Open RAN, and AI RAN Highlights,” *Dell’Oro Group* (blog), Mar. 13, 2025. [Online]. Available: https://www.delloro.com/mwc-2025-ran-open-ran-and-ai-ran-highlights/. Accessed: May 20, 2025. Dell'Oro Group

[4] ZTE Corporation, *AIR RAN White Paper*, Shenzhen, China, 2024.

[5] S. Alikhani, G. Charan, and A. Alkhateeb, “Large Wireless Model (LWM): A Foundation Model for Wireless Channels,” *arXiv* preprint arXiv:2411.08872, Apr. 7, 2025.

[6] B. Guler, G. Geraci, and H. Jafarkhani, “A Multi-Task Foundation Model for Wireless Channel Representation Using Contrastive and Masked Autoencoder Learning,” *arXiv* preprint arXiv:2505.09160, May 14, 2025.

[7] IEEE GLOBECOM 2025, “Workshop on Interplay of Foundation Models and Wireless Communications & Networks,” Call-for-Papers page, 2025.