

AI-RAN Alliance

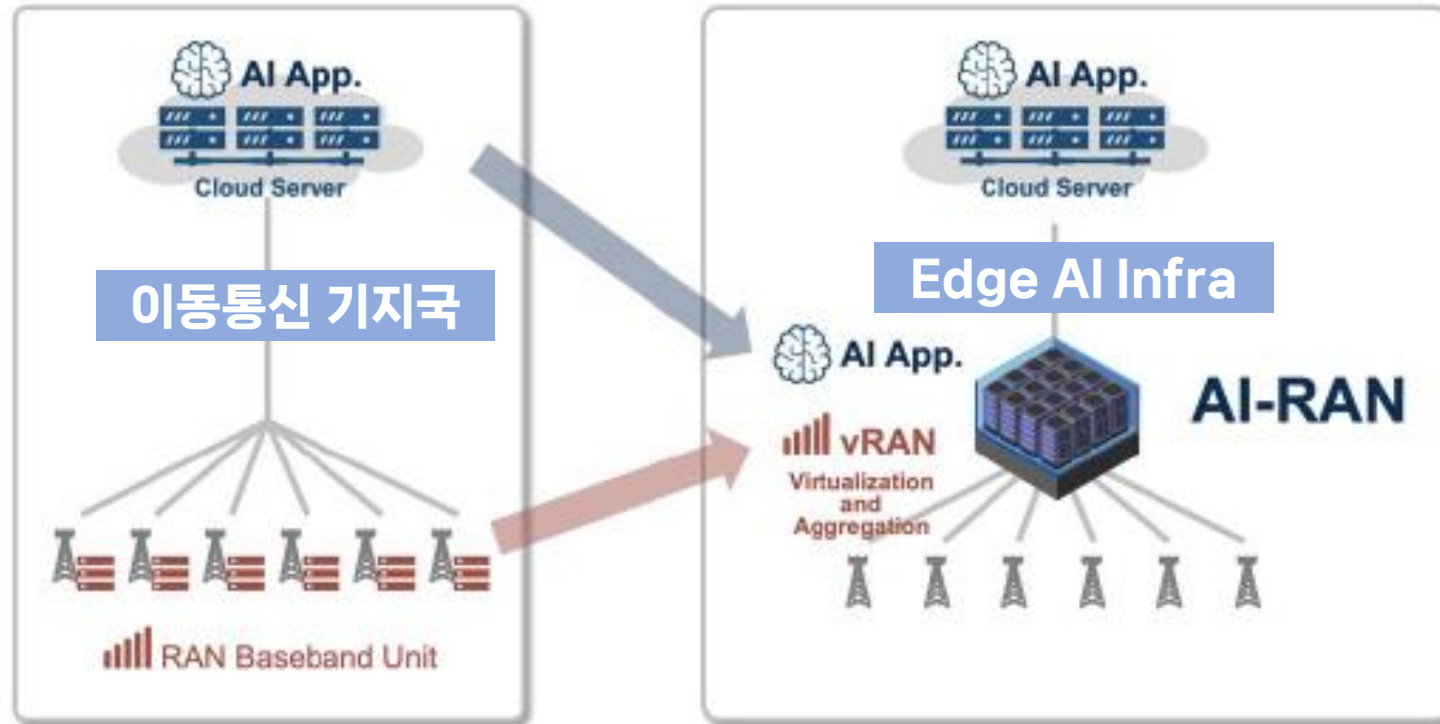


배정숙
한국전자통신연구원
2025.09.11

What is AI-RAN?

AI-RAN

- AI 기술을 무선 접속망(RAN)에 완전히 통합하여 RAN 운영 성능을 혁신적으로 향상시키고, 동일 인프라에서 AI 기반 서비스를 동시에 제공하여 신규 수익 창출 기회를 여는 기술 패러다임

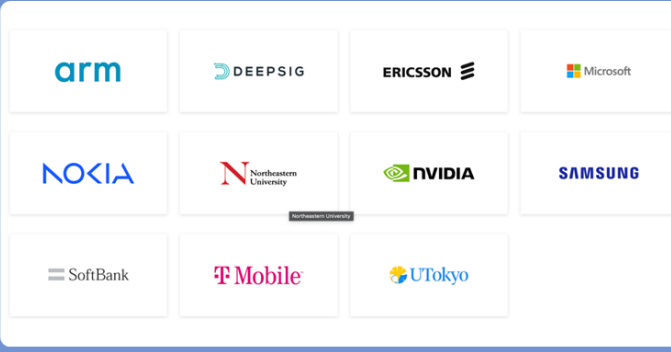


AI-RAN Alliance

AI의 잠재력을 통해 5G와 다가오는 6G 시대의 RAN을 혁신



Founding Members



General members



설립과 회원사

2024.02@MWC
11 개 창립 회원사, 98개 일반 회원사(25.08)

임무

- * Transform the network, Transform our business
- * Evolving networked sustainability

비전

- *Realizing and harnessing the potential of an AI-native RAN

AI-RAN Alliance

중점 분야

비표준화 기반의 실용 중심의 독립 연합체



Reference Design

AI 기능과 RAN 네트워크 요소 간의 통합,
무중단 운영, 성능 최적화 보장 설계
프레임워크 제시



Blueprints

AI 기능을 RAN에 효율적으로 통합하기
위한 기술 구성요소, 단계별 절차, 운영
시나리오의 체계적 정의
→ 확장 가능하고 반복 가능한 아키텍처
구현 지원



Benchmarking

다양한 RAN 환경에서의 AI/ML 모델의
성능을 정량적으로 분석 및 평가
→ 산업 전반의 요구사항과 표준화된
기대 수준에 부합하는 성능 기준 마련

AI-RAN Alliance



Committee

- Chair: Kuntal Chowdhury, NVIDIA
- Vice Chair: Dr. Felipe Arraño Scharager, Ericsson

Technical Steering Committee (TSC)

Marketing Steering Committee (MSC)

- Chair: Steph Delvoye, Head of Mobile Networks Marketing, Nokia
- Vice Chair: Yashar Nezami, Global Partnering & Ecosystem Director, Ericsson

Working Group & Task Group

WG-1: AI-for-RAN

- Chair: Chris Dick, Nvidia
- Vice Chair: Dr. Felipe Arraño Scharager, Ericsson

WG-3: AI-on-RAN

- Chair: Dr. Athul Prasad, Samsung Research America
- Vice Chair: Jihong Park(SUTD)

WG-2: AI-and-RAN

- Chair: Michele Polese, Northeastern University
- Vice Chair: Dr. Yuji Sekiya, University of Tokyo

Work Process

1

- Work item proposal at the WG
- The proposal must be as per current WG charter

2

- Work item report initiated
- Work plan defined
- Lab setup process initiated

3

- T&M Task Group defined test and benchmarking criteria, setup
- Data for AI step as necessary

4

- Lab setup and test and benchmarking executed
- Results discussed at WG
- TSC review of the results approval

5

- Board approval for publication
- WI report publish (Dataset, and any Notebook if
- Optionally Blueprint published

WG-1: AI-for-RAN

Enhancing network performance

AI-for-RAN

AI for the
enhancement of
RAN



Spectral Efficiency

목표

- AI를 RAN에 적용하여 성능 향상을 실현하는 기술을 탐색하고 발전시키는 것

추진 분야

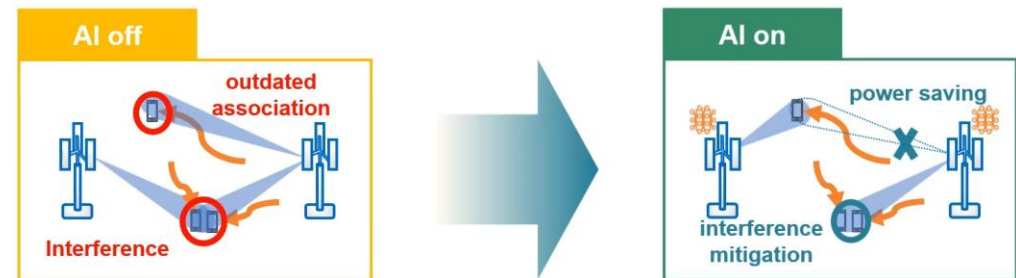
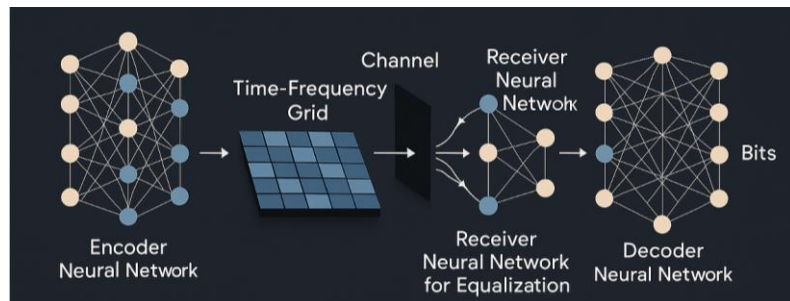
- RAN의 효율성, 용량 및 기타 KPI를 향상시키기 위한 AI-네이티브 RAN 솔루션을 발굴/정의/구축 (3GPP 및 기타 산업 단체의 AI/ML 활동을 활용 및 보완)

핵심 기술 영역

- Radio Link Optimization: AI 기반 신호 처리 기술을 통해 링크 품질 및 안정성 개선
- Spectral Efficiency: 주파수 자원 활용을 극대화하며 고품질 사용자 경험을 유지
- Performance: 고신뢰성, 고용량, 광역 커버리지, 고에너지 효율 (혼잡하거나 요구사항이 높은 환경에서도 QoS 유지 가능)

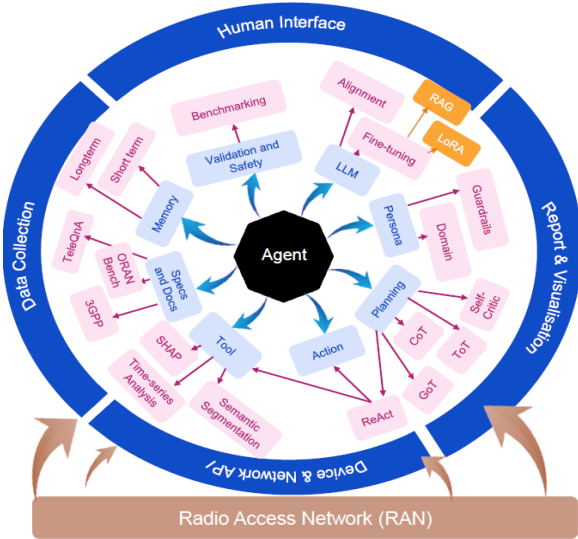
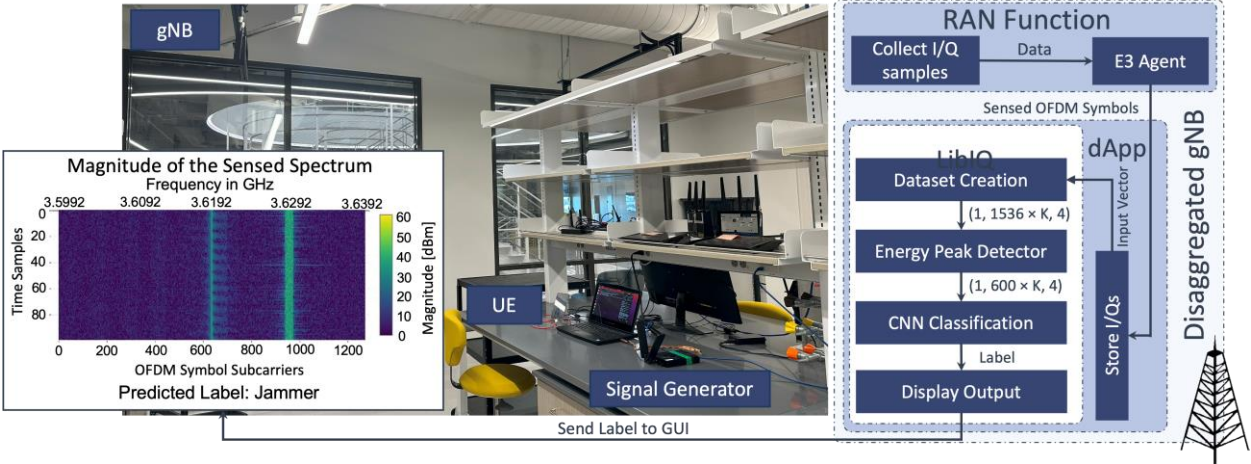
WG-1 Work Items

WI#	Title	목표
1	Learned Air Interface with Online Learning (DeepSig, NVIDIA)	파일럿 없는 Autoencoder 기반 PHY 설계, 빠르게 변하는 채널에서도 안정적 전송, DMRS+QAM 대비 Throughput 최대 70% ↑, BLER 개선, Overhead 감
3	AI-based PUSCH Channel Estimation (Keysight)	다양한 시나리오에서 PUSCH 채널 추정 성능 향상, 셀 엣지 사용자 품질 개선(전송량 30% ↑)
5	AI-based 5G Beamforming for Mobility-Aware Interference Mitigation and Power Saving (Yonsei, Viavi, SUTD)	모빌리티 감지 간섭 완화 + 전력 절감 throughput ↑, power consumption ↓



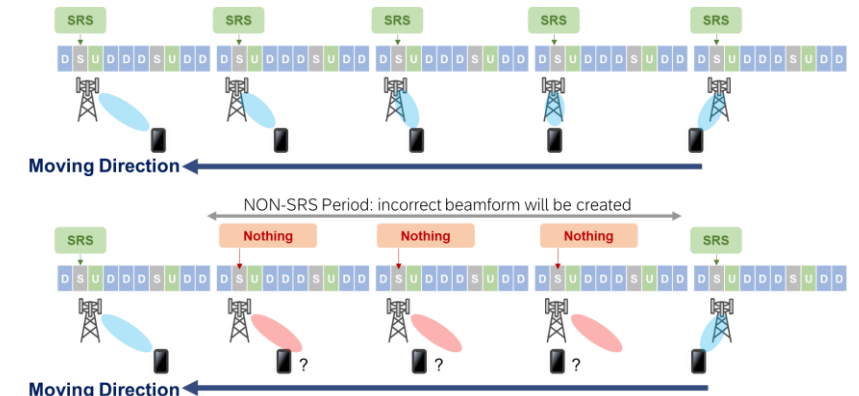
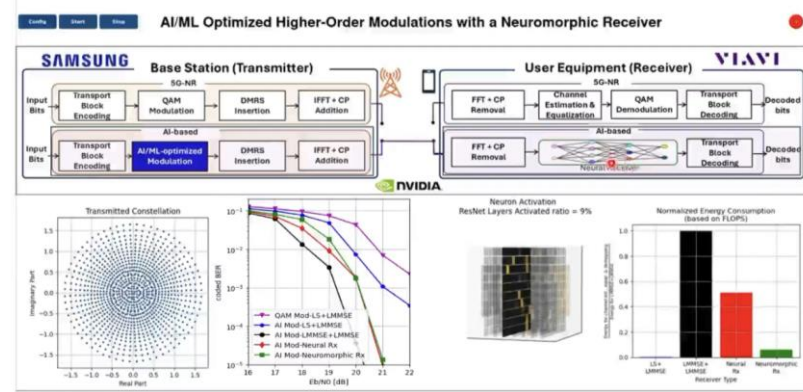
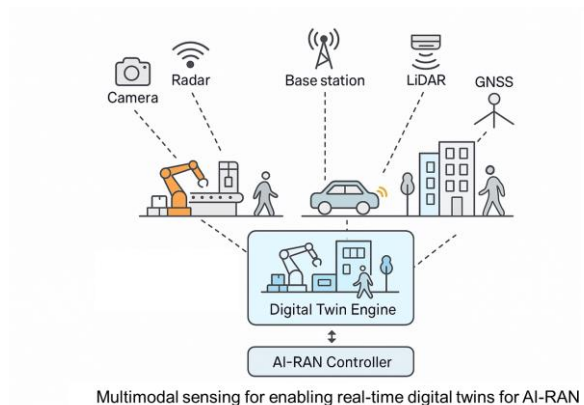
WG-1 Work Items

WI#	Title	목표
6	AI-based Spectrum Sensing in the RAN (NEU)	AI 기반 동적 스펙트럼 감지 및 간섭 탐지 최적화
7	Cell energy saving (Ericsson, Keysight)	AI로 트래픽 기반 셀 on/off 최적화, 에너지 절감
8	Agentic Architecture for AI-native RAN (U Leeds, NVIDIA)	Agentic multi-agent RAN 운영 구조 확립



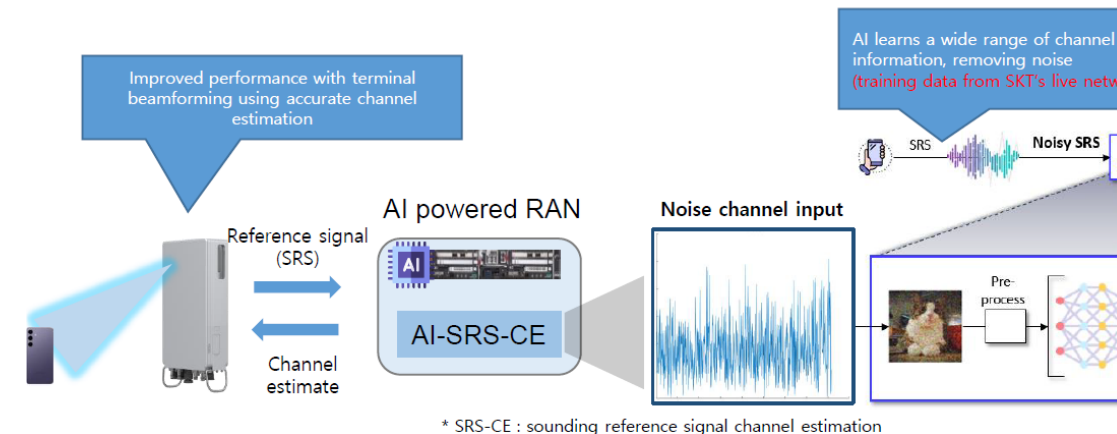
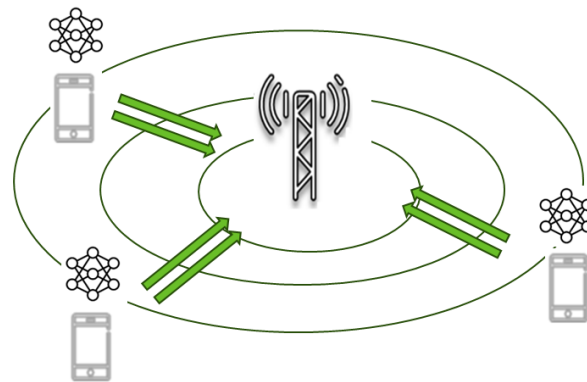
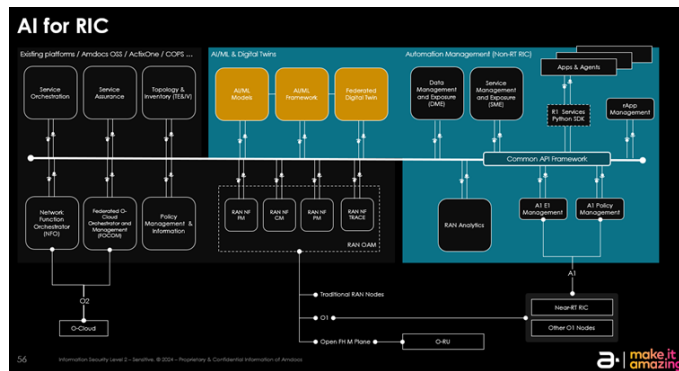
WG-1 Work Items

WI#	Title	목표
9	Multimodal sensing (U Oulu, NVIDIA)	CSI + 다중 센싱(USRP+LiDAR+영상) 융합으로 beam selection 최적화
10	Neuromorphic transceiver (VIAVI)	Neuromorphic 기반 PHY 설계(CE, 보상, 심볼 매핑).로 저전력·고효율 구현
11	AI for resource allocation (NEU)	AI 기반 RAN slice QoS 보장 및 SLA 관리
13	SRS prediction (SoftBank, NVIDIA)	AI 기반 SRS/CSI 보간(SRS 전송 간격 보완 → CSI 예측 정확도 ↑, UX 개선)으로 overhead 줄이고 성능 개선



WG-1 Work Items

WI#	Title	목표
14	AI-for-RIC (Amdocs)	RIC에 최적화된 AI 기능 아키텍처 개발
15	AI Tx Diversity & Lifecycle Management (MediaTek, Keysight)	Tx Diversity 강화 + AI 모델 생애주기 관리
16	AI SRS Channel Estimation (SKT, Samsung)	AI 기반 SRS 추정, 상용망 데이터 활용 AI 기반 CE → cell-edge DL throughput +10%
17	AI-Assisted Ghost Preamble Detection (GlobalLogic)	Ghost preamble 검출 오류 개선 → 자원 낭비 ·coverage 저하 방지



WG-2: AI-and-RAN

Shared Infrastructure

AI-and-RAN

AI and RAN
sharing the same
infrastructure



Asset Utilization

목표

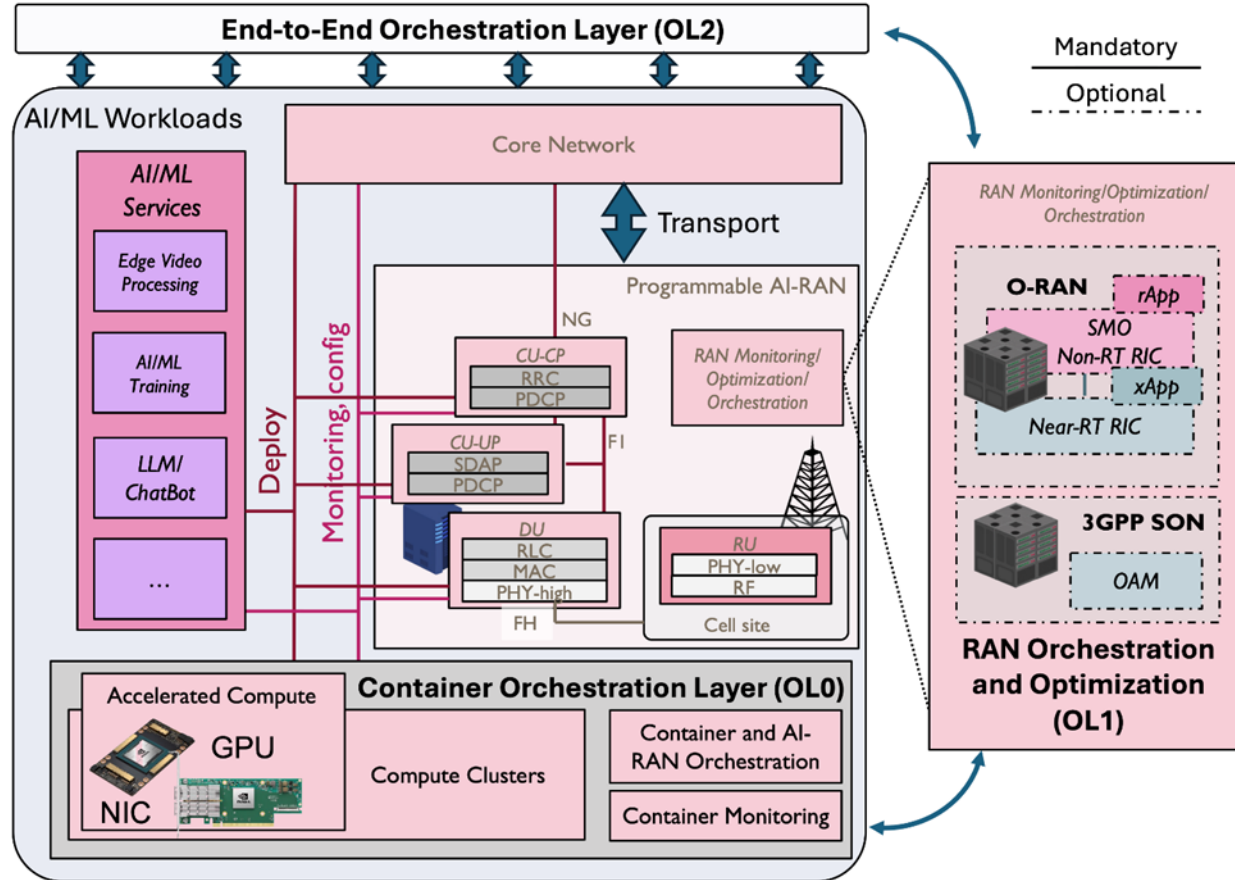
- AI/GenAI 응용과 RAN 기능을 하나의 통합 인프라에서 동시에 실행하기 위한 기술 연구
- 컴퓨팅 자원의 효율적 활용 및 AI 서비스 수익화 기회 확보

핵심 기술 영역

- 통신+컴퓨팅 융합 인프라 활용 모델 연구: RAN과 AI/GenAI 워크로드의 병렬 실행을 통한 자원 활용률 극대화, 특히 멀티테넌시 환경에서 리소스 분할·공유 기술을 단계별로 실증(사전 구성 기반 정적 분할, 규칙 기반 동적 분할, AI 기반 동적 자원 관리)
- RAN 가상화와 엣지 집약형 처리 모델 연구: 가상화된 RAN (DU/CU/UPF 등)을 엣지에 배치, 통계적 트래픽 평균화를 통한 인프라 최적 활용

WG-2: AI-and-RAN

AI-RAN Infrastructure



* 배치 모드: 전용 서버 → 공유 리소스 파티셔닝 → 완전 공유

소프트웨어 구성

- OL0: 컨테이너 오케스트레이션 계층 (예: Kubernetes)
- OL1: RAN 오케스트레이션 (SON, RIC, SMO)
- OL2: End-to-End 오케스트레이션 (정책/워크플로우 통합 관리)

하드웨어 구성

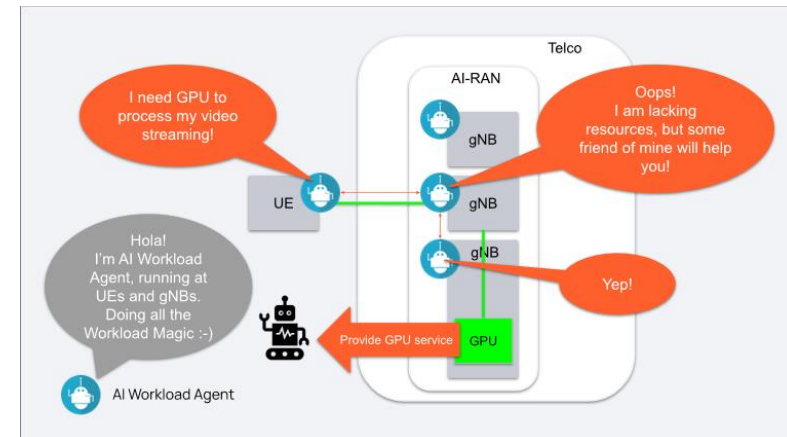
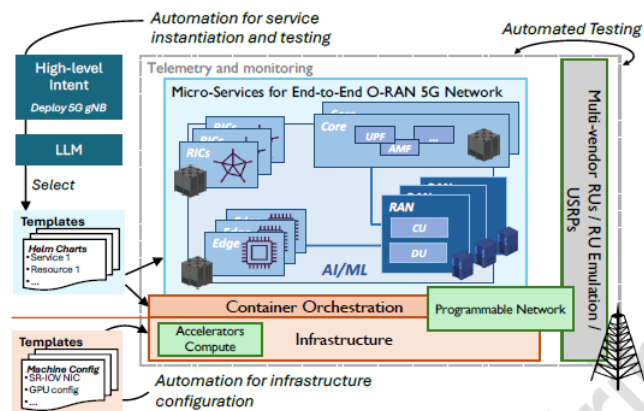
- 고성능 CPU, GPU, FPGA, TPU, DPU 등 고대역폭 메모리 및 고속 스토리지 (랜덤 액세스 최적화)
- SDN 기반 네트워크 인터페이스 (시간 동기화 포함)

주요 기술

- 컨테이너화, Kubernetes 중심 인프라

WG-2 Work Items

WI#	Title	목표
1	Evaluation of AI-RAN Infrastructure Methodologies (Univ. Tokyo)	AI-RAN 인프라 오케스트레이션 방법론 평가 및 가이드라인 마련. 리소스 할당/분리, 모니터링 기법 비교 및 최적 조합 제시
2	Intelligent and Open Architecture for AI-RAN (Northeastern, zTouch)	차세대 사용자 인터페이스(zTouch)를 RAN과 AI-native O-RAN 환경에서 통합 검증, 사람의 촉각·제스처 인식 → RAN 제어/서비스 연동 실험.
3	Decentralized Management of Telecom Networks with Swarm Intelligence (GlobalLogic)	Swarm Intelligence (집단 지능) 기반 RAN 제어 방안 연구, 분산된 에이전트들이 협력적 의사결정으로 네트워크 최적화 달성. 방법



WG-3: AI-on-RAN

AI-on-RAN WG(Enabling Edge Services)

AI-on-RAN

AI applications
enabled by RAN



New Applications

목표

- AI 및 생성형 AI(GenAI) 응용을 RAN 상에서 실행하기 위한 무선 인터페이스 요구사항 정의
- 5G 환경에서 응용 성능 벤치마크 → 향후 6G 시스템에 필요한 신규 요구사항 도출
- 지연(Latency), 처리량(Throughput), 지터(Jitter), 패킷 지연(Packet Delay), 암호화 요구사항

핵심 기술 영역

- 현재의 AI/ML 및 GenAI 기술 검토, 기술적 난제를 파악하며, 활용 사례를 정의하고, 성능 테스트용 랩 시스템 개발

WG-3: AI-and-RAN

Literature and Standards Survey

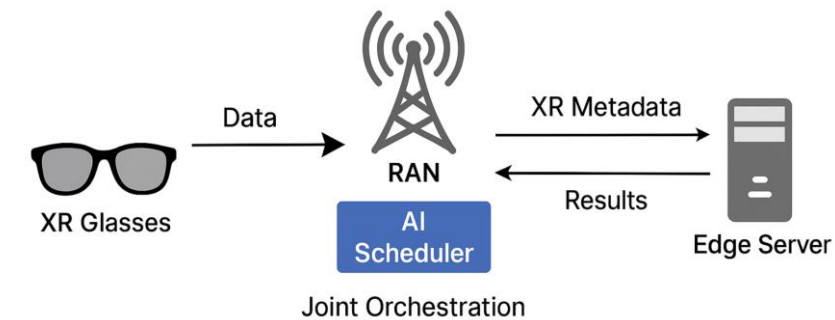
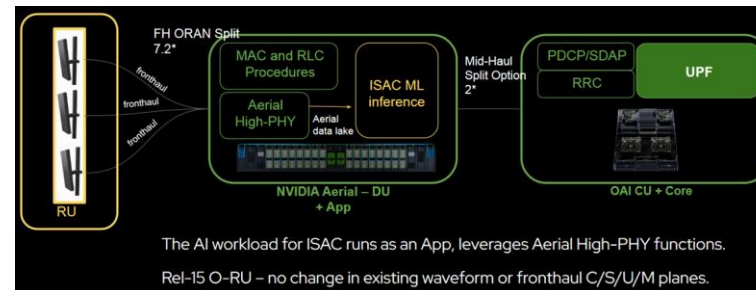
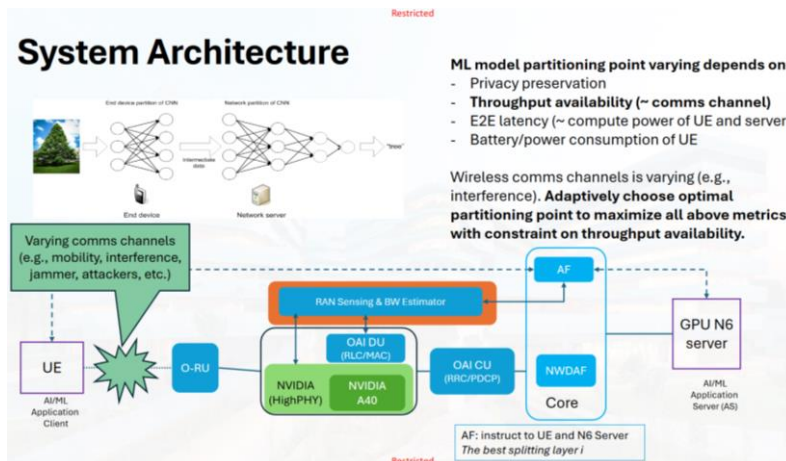
- AI 및 GenAI의 RAN 적용을 위한 표준(3GPP, O-RAN), 기술 발전, 아키텍처 및 요구사항 조사
- LLM, VLM, AI Agent 기술의 확산에 따라, 이를 RAN에 적용하기 위한 기술적 고려 및 요구사항 검토
- 6G 엣지 자원의 효율적 활용을 위한 API, KPI, 과금·보안 체계 등 엔드 투 엔드 요건 제시

Use Cases, Requirements, and Problem Statements

- AI-on-RAN 실현을 위한 RAN 요구사항, 유스케이스, 기술 과제 정의
(MEC/클라우드 기반 실행 구조 중심)
- 4가지 사용 사례 : 시각 분석, 예측 기반 통신, 다중 에이전트 협력, 생성형 엣지 AI
- 기술 요구사항: 초저지연($\leq 5\text{ms}$), 고처리량($> 1\text{Gbps}$), 낮은 지터, 고성능 연산(GPU/VRAM), 강화된 보안·프라이버시
- 핵심 과제: 종단간 지연 보장, 추론 파이프라인 최적화, 모델/단말 이기종성, 신뢰성 확보

WG-3 Work Items

WI#	Title	목표
1	AI enabled XR with split compute (SUTD)	차세대 네트워크(6G+)에서 AI 기반 애플리케이션(특히 Intelligent Video Analytics, IVA)을 지원하기 위한 Split Compute 구조 연구, XR 서비스에서 latency/에너지/privacy 개선
2	AI enabled critical applications (Tiami)	6G RAN 환경에서 Mission-Critical Applications (MCA) (예: 원격 제어, 산업 자동화, 긴급 통신)을 위한 AI 지원 기술 검토.
4	Cellular XR/AI Glasses with AI-on-RAN (NEU)	셀룰러 연결 없는 XR/AI 글래스를 위해, RAN 인프라에 AI를 통합해 통신과 추론을 공동 최적화하는 지능형 오프로딩 구조를 구축



Task Group: Data for AI

- 목적

- (1) AI 학습 및 추론을 위한 고품질 데이터 수집 방법론 정립
- (2) 실험실 환경(예, Keysight, NEU, SUTD 등)에서의 데이터 수집 및 공유 프레임워크 설계
- (3) 디지털 트윈 기반 시뮬레이션 + 실데이터 연계를 통한 AI 성능 향상 목표

- 실환경 실시간 데이터 문제점

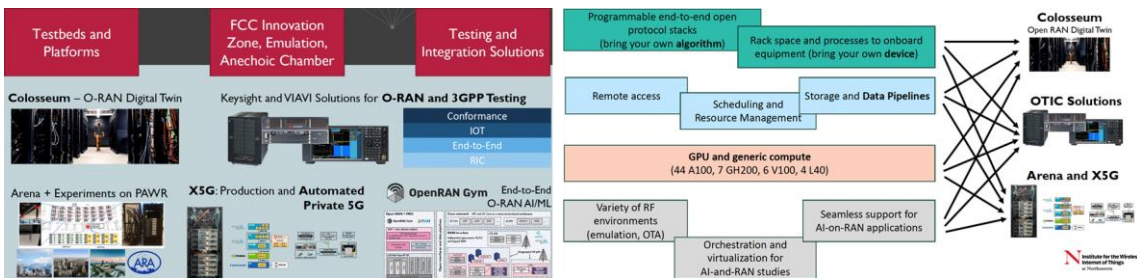
- 실 네트워크 데이터는 프라이버시, 소유권 문제로 공유 어려움
- 실시간 제약으로 세밀한 데이터 수집 불가
- 실환경에서 발생 가능한 이상 상황의 데이터 확보 어려움

- D4AI 해결 방향

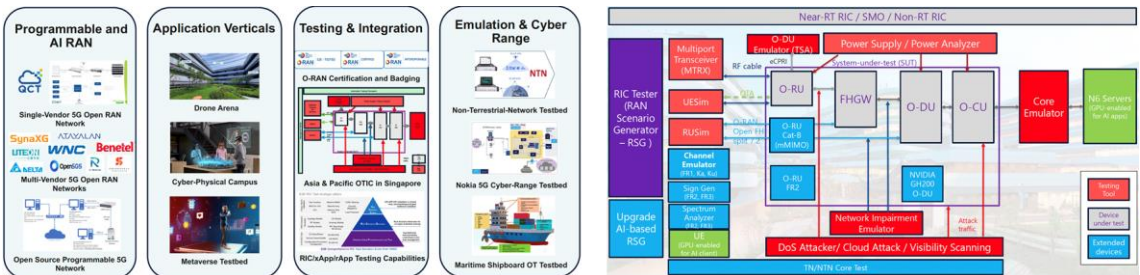
- 실험실 기반 + 디지털 트윈 기반 데이터셋 병행 수집
- Synthetic 데이터 생성 허용
- 다양한 계층/인터페이스에서 세밀한 수집 가능한 구조 설계

AI-RAN Alliance Lab Proposal

[NEU] Open6G Solutions for the AI-RAN Lab

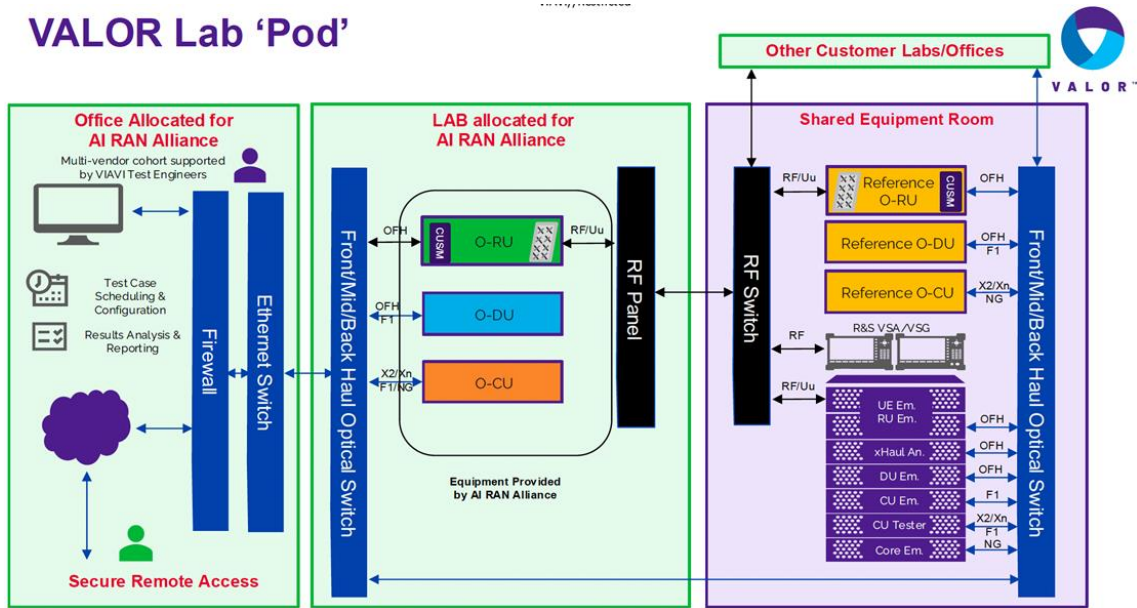


[SUTD] Future Communications Connectivity (FCCLab)

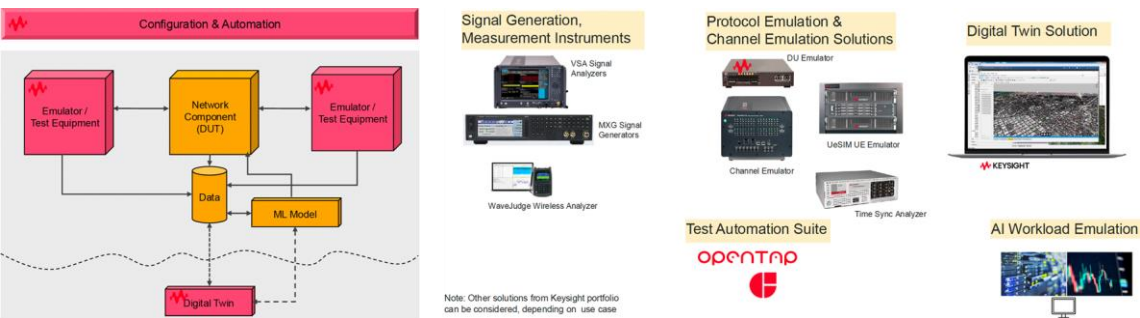


[VIAVI] VIAVI Automated Lab-as-a-Service for Open RAN (VALOR)

VALOR Lab 'Pod'



[Keysight] AI Lab.



Summary

- **왜 중요한가?**

기존 RAN의 한계 극복: 고비용, 낮은 유연성, 수익성 저하
AI와 RAN의 융합: 실시간성, 자율성(Self-X), 고효율
운용

Edge 중심의 AI 서비스 실현: 자율주행, 생성형 AI 등

- **핵심 효과**

CapEx / OpEx 절감 → 투자 회수 극대화

AI 기반 신규 서비스 수익화 (AI 슬라이싱, inference 등)

에너지 효율 및 네트워크 적응성 향상

- **생태계 협력 가속화**

AI-RAN Alliance 중심으로 글로벌 확산

MWC 2025 데모 및 WI 연구를 통해 기술 실현성 검증

디지털 트윈 + AI-native 아키텍처 실증 본격화

- **지금이 골든타임**

AI-RAN은 6G 및 초지능형 서비스의 핵심 인프라

산업·학계·연구기관이 함께 만드는 개방형 협력 생태계
필요

감사합니다.
Thank you.

JungSook Bae(jsbae@etri.re.kr)

