

Proof of AI-RAN Concept

Jenq-Shiou Leu

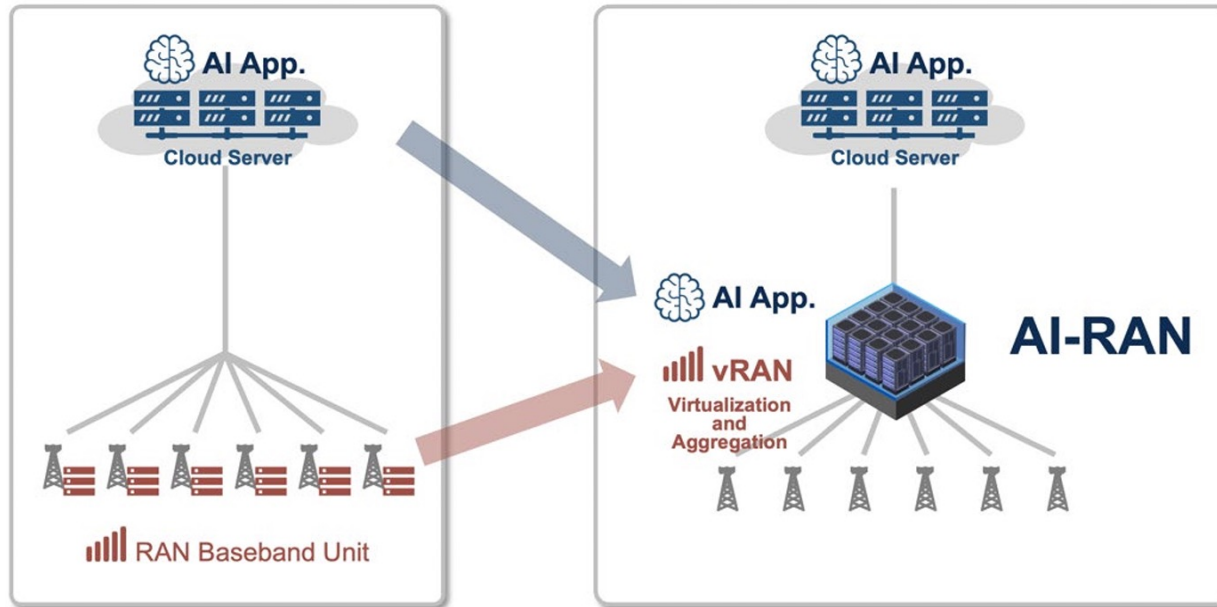
Dean, College of Electrical Engineering and Computer Science (EECS), NTUST

Professor, Department of Electronic and Computer Engineering (ECE), NTUST

“Open Integrated Optical and Wireless Prototype System” supported by NSTC, Taiwan

AI-RAN: AI-Radio Access Networks

AI-RAN integrates AI capabilities into RAN operations over a common accelerated infrastructure.



AI-RAN Alliance

Three Items	AI-for-RAN	AI-and-RAN	AI-on-RAN
			
Focus Areas	Focuses on using AI to optimize the performance of the RAN itself	Focuses on the coexistence, synergy, and management of AI and RAN workloads on the same infrastructure	Explores utilizing the RAN infrastructure to provide external AI services and applications
Applications	• AI-Native air interface and signal processing • Positioning and beam management • Radio resource management and scheduling • Energy and spectrum efficiency • Network optimization and anomaly detection	• Design architecture and components of multi-tenant system • Lifecycle management of AI and RAN workloads • Validating the multitancy scenarios • Data center optimization for AI and RAN workloads	• AI-based multimedia applications • AI-based security • AI-based automation and industrial applications • GenAI/AI-based network services • Efficient AI/ML model splitting

AI/ML Platform

AI/ML Training Platform

- Model Training
- Model Fine-Tuning
- Model Management

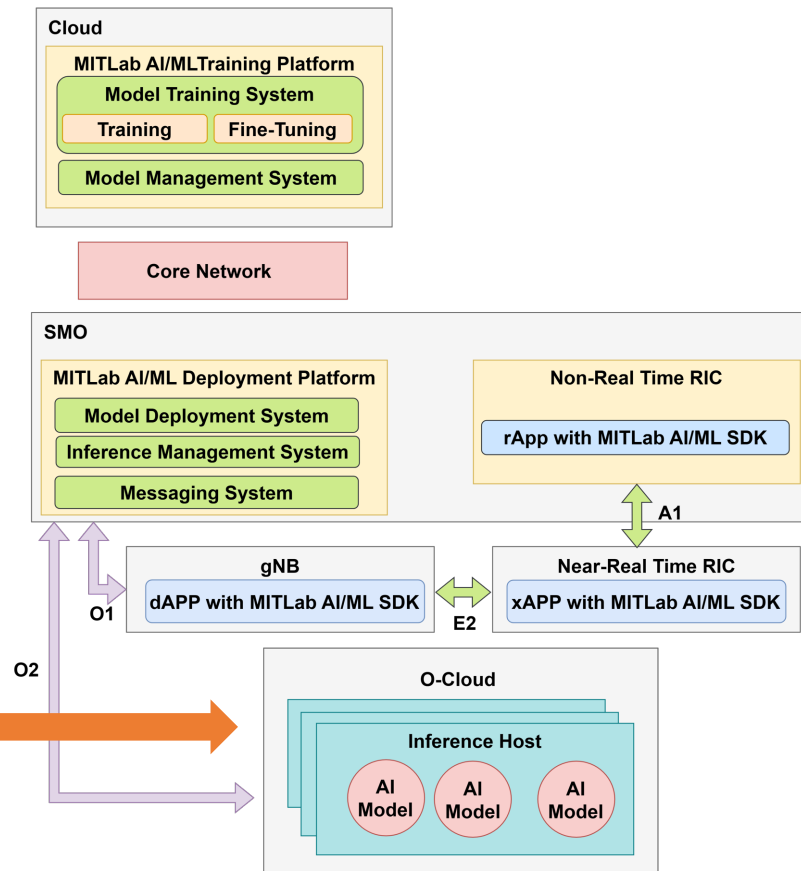
AI/ML Deployment Platform

- Model Deployment
- Metric & Result Collection
- Inference Node Real-Time Monitoring

Inference Host

- Model Inference Service
- AI Service Load Balancing

AI-and-RAN

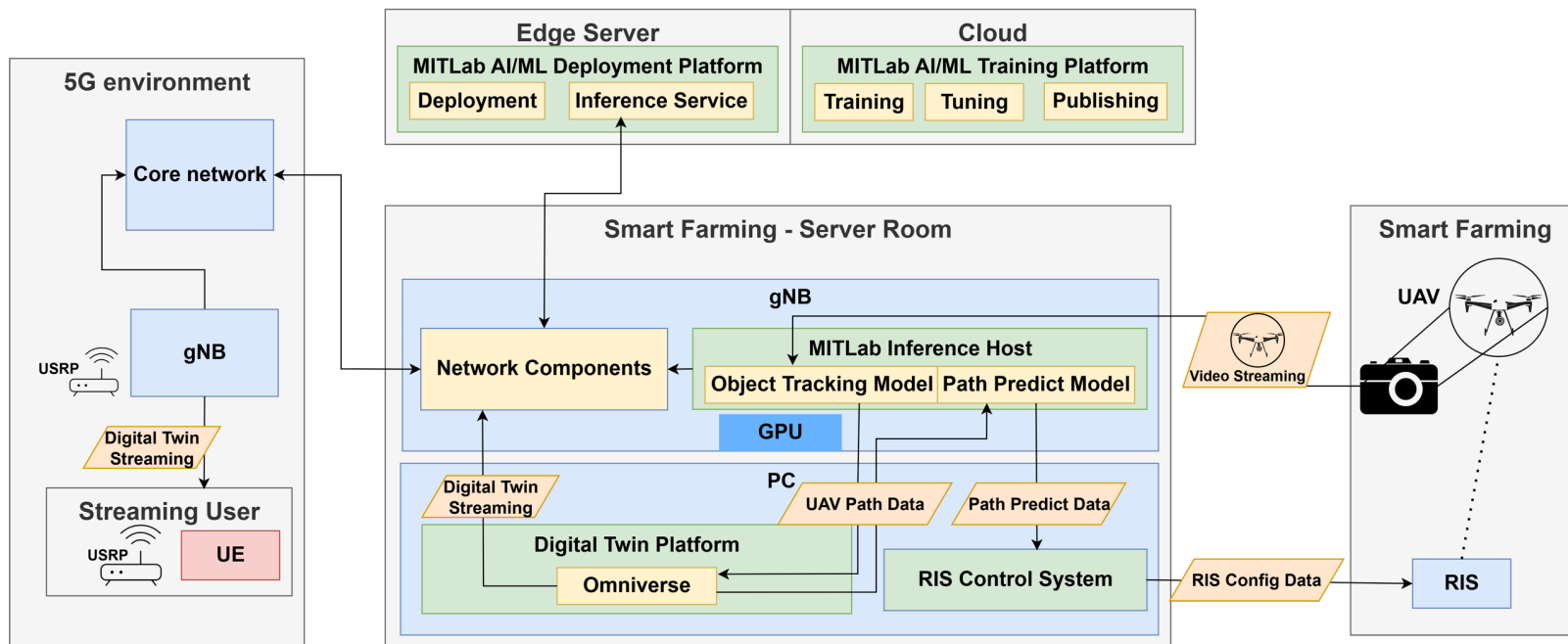


Use Case: Smart Farming

AI-on-RAN

- AI/ML Training Platform
- AI/ML Deployment Platform

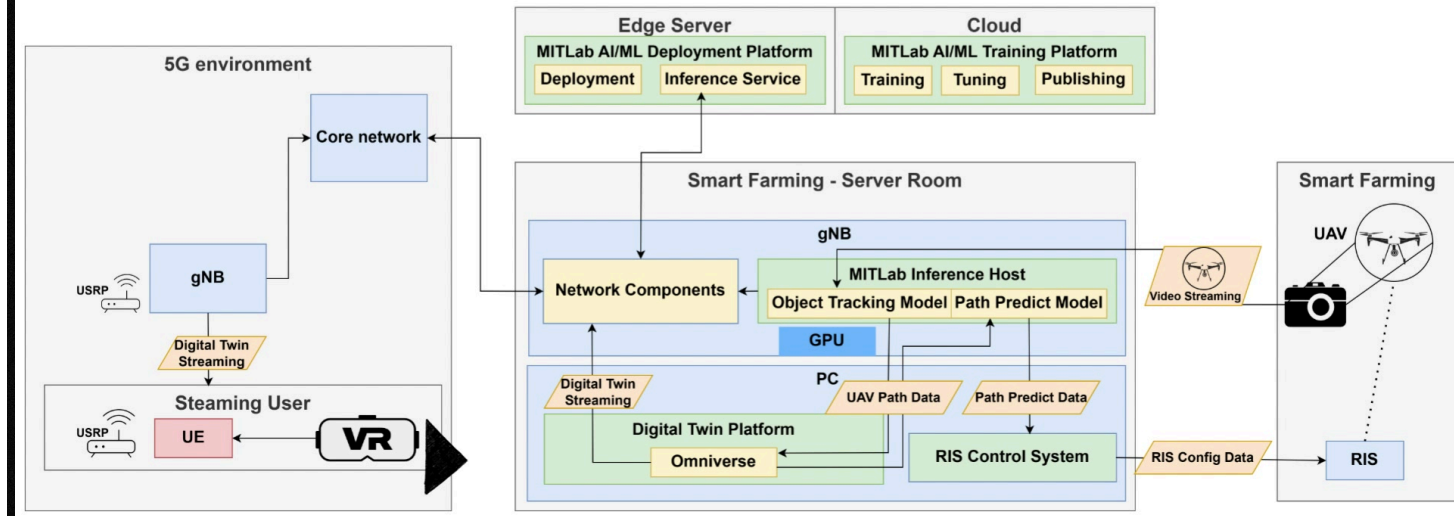
- Digital Twin Platform
- Tracking System



Tracking System & Digital Twin Integration

Smart Farming Usecase

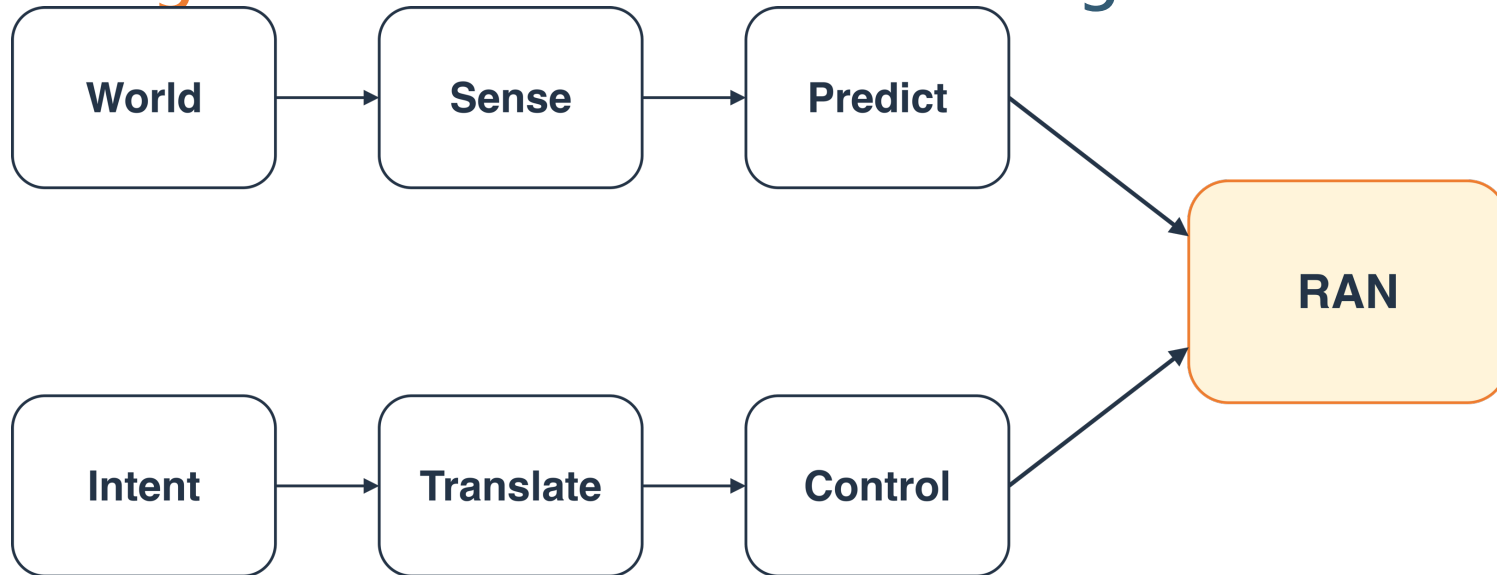
- Digital Twin Platform
- AI-Tracking System



Intent-based Network

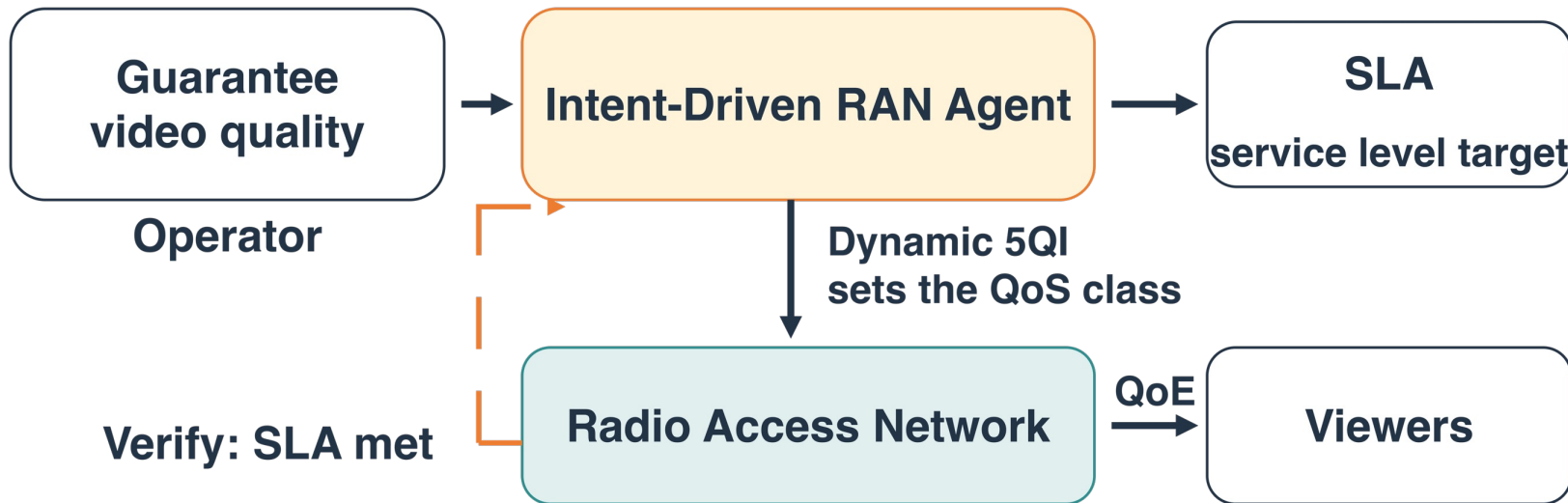
AI-for-RAN

Sensing and **Intent** Drive the RAN Together



One Sentence Becomes a Service Level

Plain words in, **measurable targets out**



QoS / QoE: Quality of Service / Quality of Experience SLA: Service Level Agreement

The RAN Agent Running on a Live Testbed

Energy saving, approved by a human

Three things to watch for

- 1 Intent
- 2 Agent action
- 3 Network result

The screenshot displays a live testbed environment. On the left, a terminal window shows logs from the RAN Agent, including occupancy samples, PEE state transitions (active to sleep), and UPF data. A red vertical line separates the logs from the iPerf3 interface on the right. The iPerf3 interface shows a server address of 192.168.137.79, server port 5201, and a stream of 5. The 'Start' button is highlighted. Below the iPerf3 interface, a text box explains the setup: 'Use ADB and FFmpeg to remotely control the UE and run iPerf, creating a high-traffic scenario. ADB is the Android Debug Bridge, UE is user equipment, and iPerf measures network throughput.'

使用 ADB 與 FFmpeg 遠端操控 UE，並執行 iPerf 模擬大流量情境。ADB 是 Android Debug Bridge，UE 是使用者終端設備，iPerf 用來量測網路吞吐量。

Use ADB and FFmpeg to remotely control the UE and run iPerf, creating a high-traffic scenario. ADB is the Android Debug Bridge, UE is user equipment, and iPerf measures network throughput.

What This Work Contributes

- 1 AI/ML platform and digital twin
- 2 Intent-based network
- 3 RAN agent on a live testbed

Open to EU collaboration

- Shared intent formats
- Benchmarks for RAN agents

Prof. Jenq-Shiou Leu

Dean, College of Electrical Engineering and Computer Science,
National Taiwan University of Science and Technology

Contact

