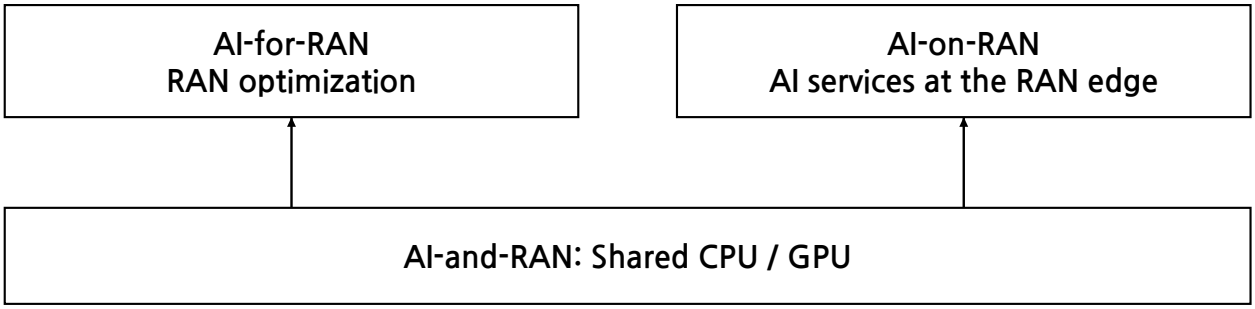
AI Radio Access Network (AI-RAN)

1. Overview

Definition	A **networking technology** that combines AI-based optimization and computing infrastructure with the radio access network **to improve communication efficiency and deliver AI services.**

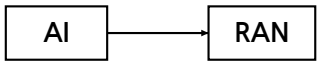
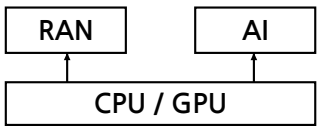
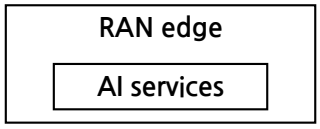
2. Concept and perspectives

A. Concept diagram



The three perspectives can be applied together.

B. Three perspectives

Category	Diagram	Description
AI-for-RAN		**RAN optimization** AI for channel estimation, radio resource allocation and energy control.
AI-and-RAN		**Shared computing** Run RAN and AI workloads on shared CPU and GPU resources.
AI-on-RAN		**AI service delivery** Deliver AI services with inference and low-latency connectivity at the RAN edge.

3. Key technologies

Category	Technology	Description
Computing	**Hardware acceleration**	GPU and other accelerators for AI inference and signal processing.
	Network function virtualization	Separate RAN functions from hardware and run them as software.
Operations / control	**Resource orchestration**	RAN / AI workload placement, scaling and compute resource allocation.
	RAN Intelligent Controller (RIC)	Optimize RAN through policy and radio resource control (optional).
	Model lifecycle management	Data collection → training and validation → deployment → inference and monitoring

Design requirements

- **Real-time performance**: Meet radio processing deadlines and control resource contention.
- **Security**: Isolate workloads, control access and protect data.