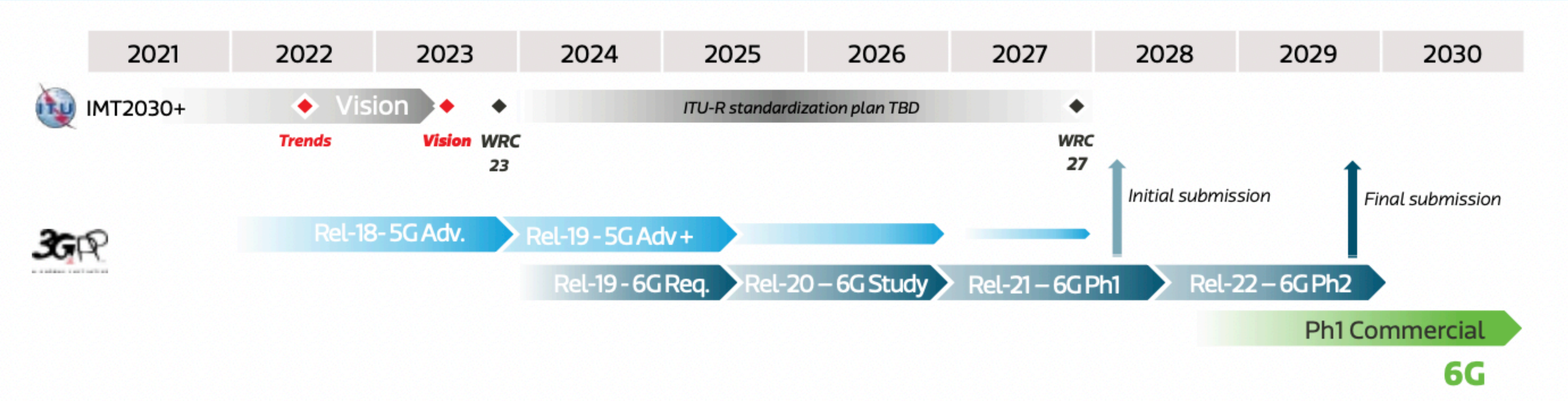
AI and 6G: Opportunities and Challenges

Satish Kumar, Airspan Network

6G Timeline



[MediaTek 6G vision white paper]() (2022)

6G Vision

Connected machine - machine as a main user

AI - new tool for communication

Openness in mobile communication

Social goal

[Bharat 6G Vision]

[University of Oulu 6G vision]

[University of Bristol (6G Futures)]

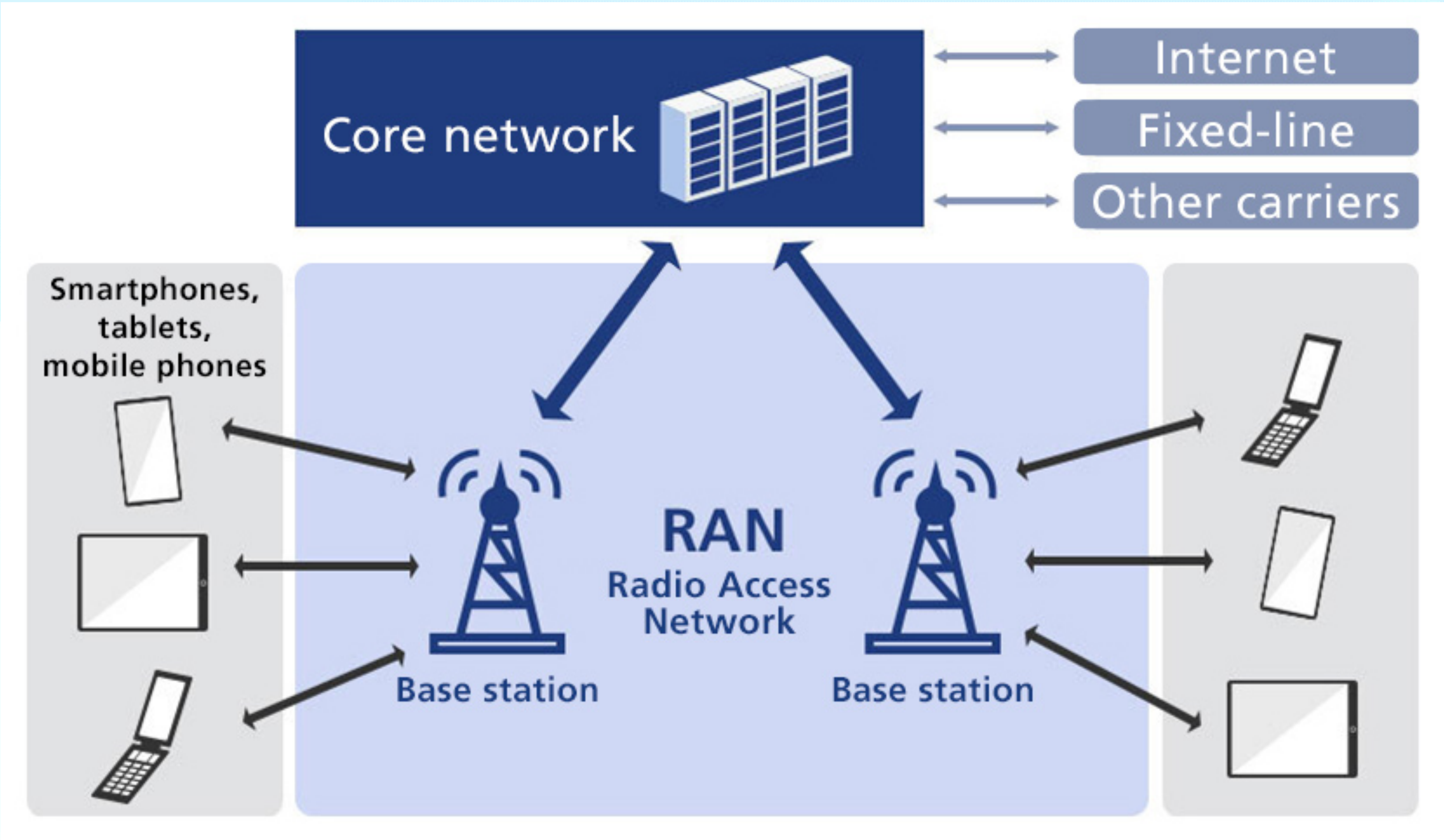
[NGMN 6G Drivers and vision]

[Europe 6G vision]

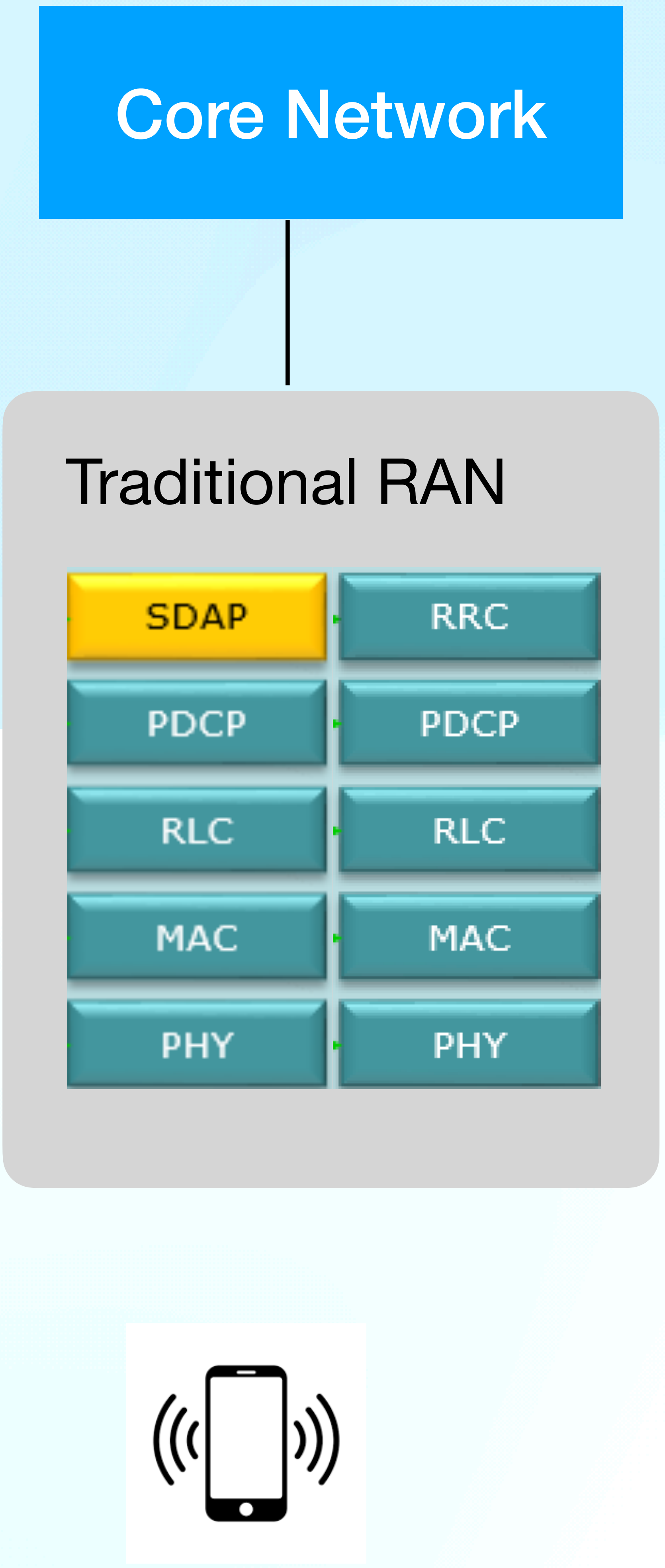
[Next G Alliance (ATIS) Vision]

[Samsung 6G Vision]

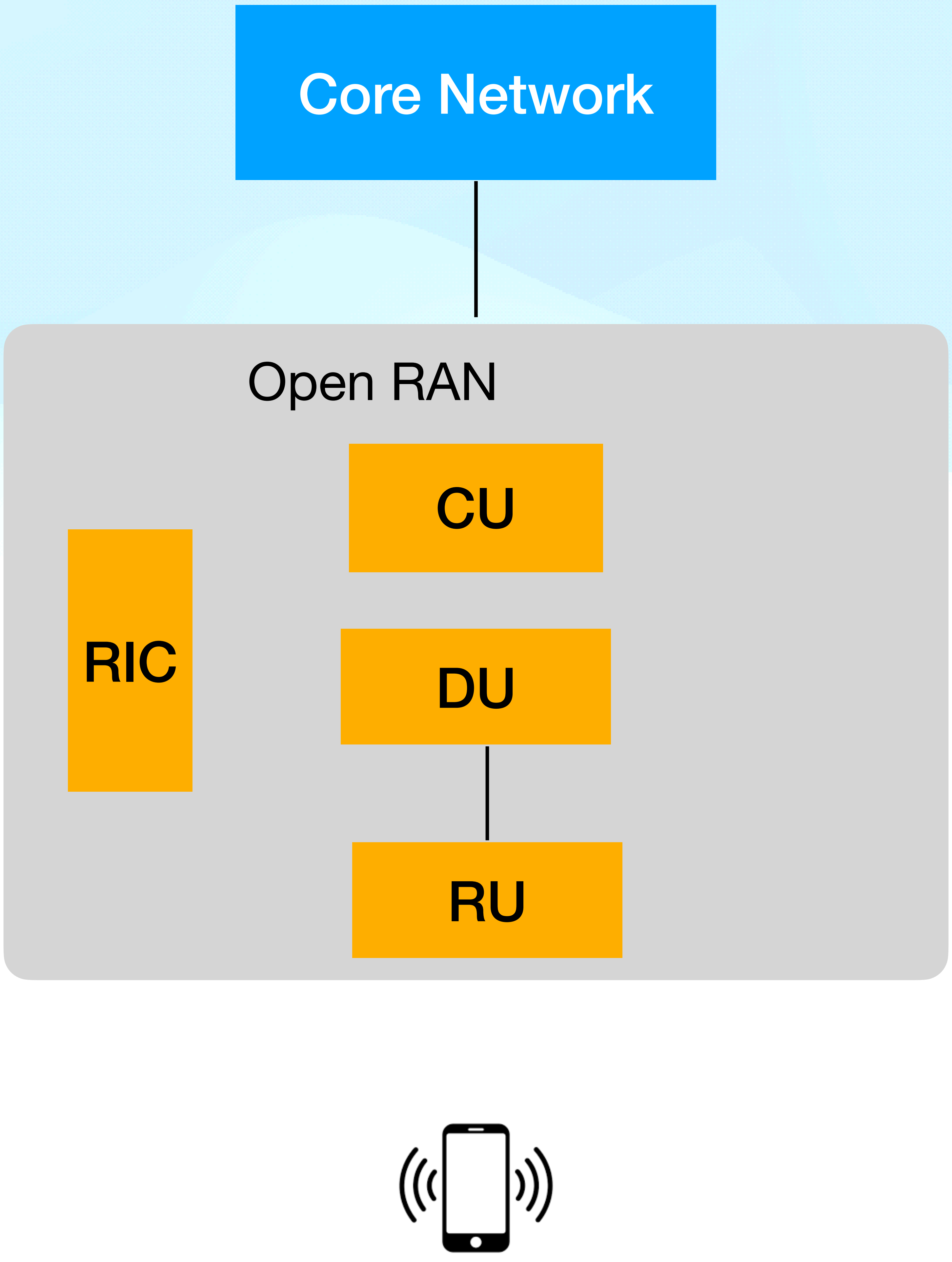
End -to-end 5G Network



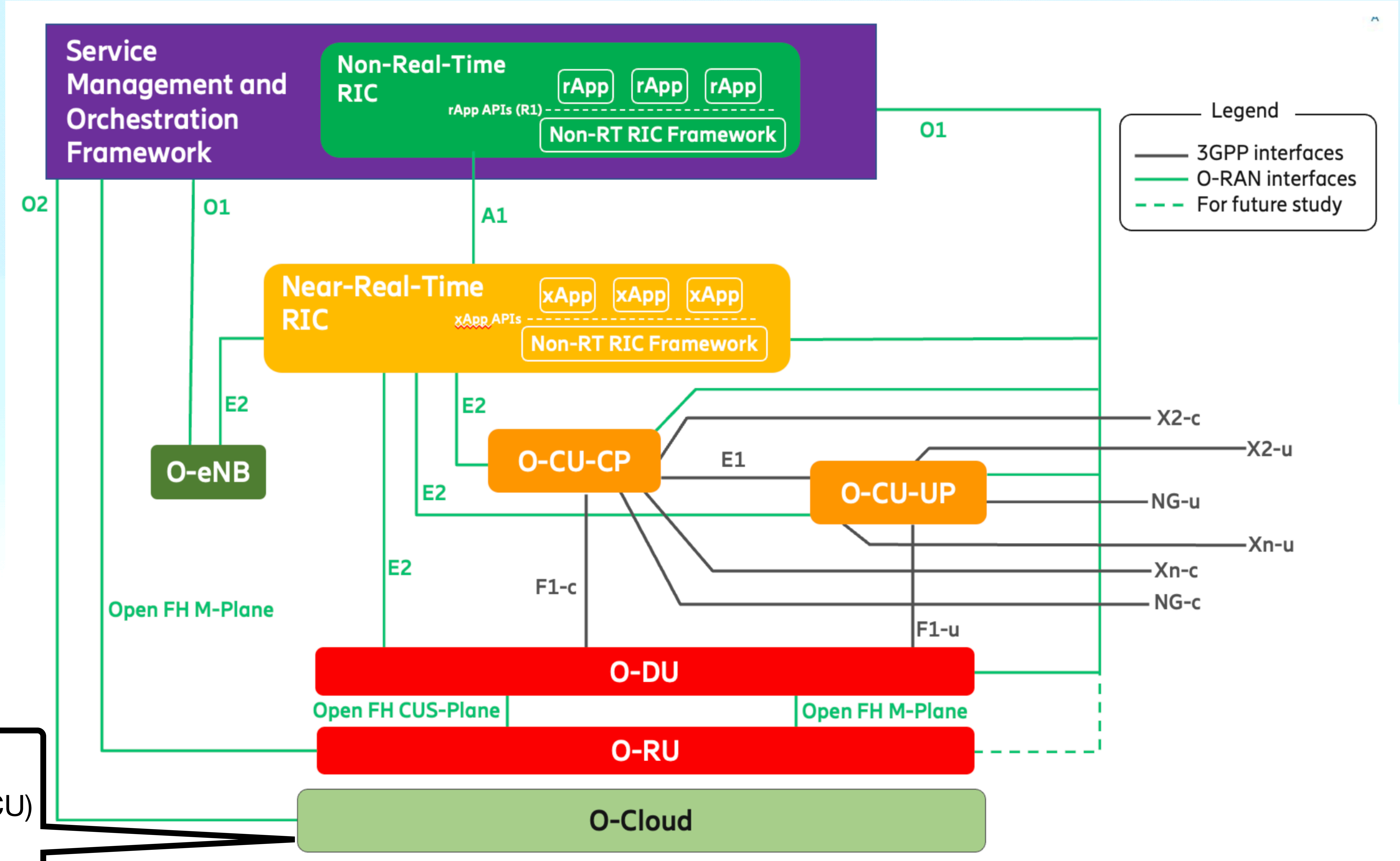
Traditional RAN vs Open RAN



- Disaggregation
- Decoupling HW from SW
- Open / enlarge vendor ecosystem
- Open interfaces
- Intelligent management
- Lower CAPEX
- Minimisation of proprietary solution



Open RAN Architecture



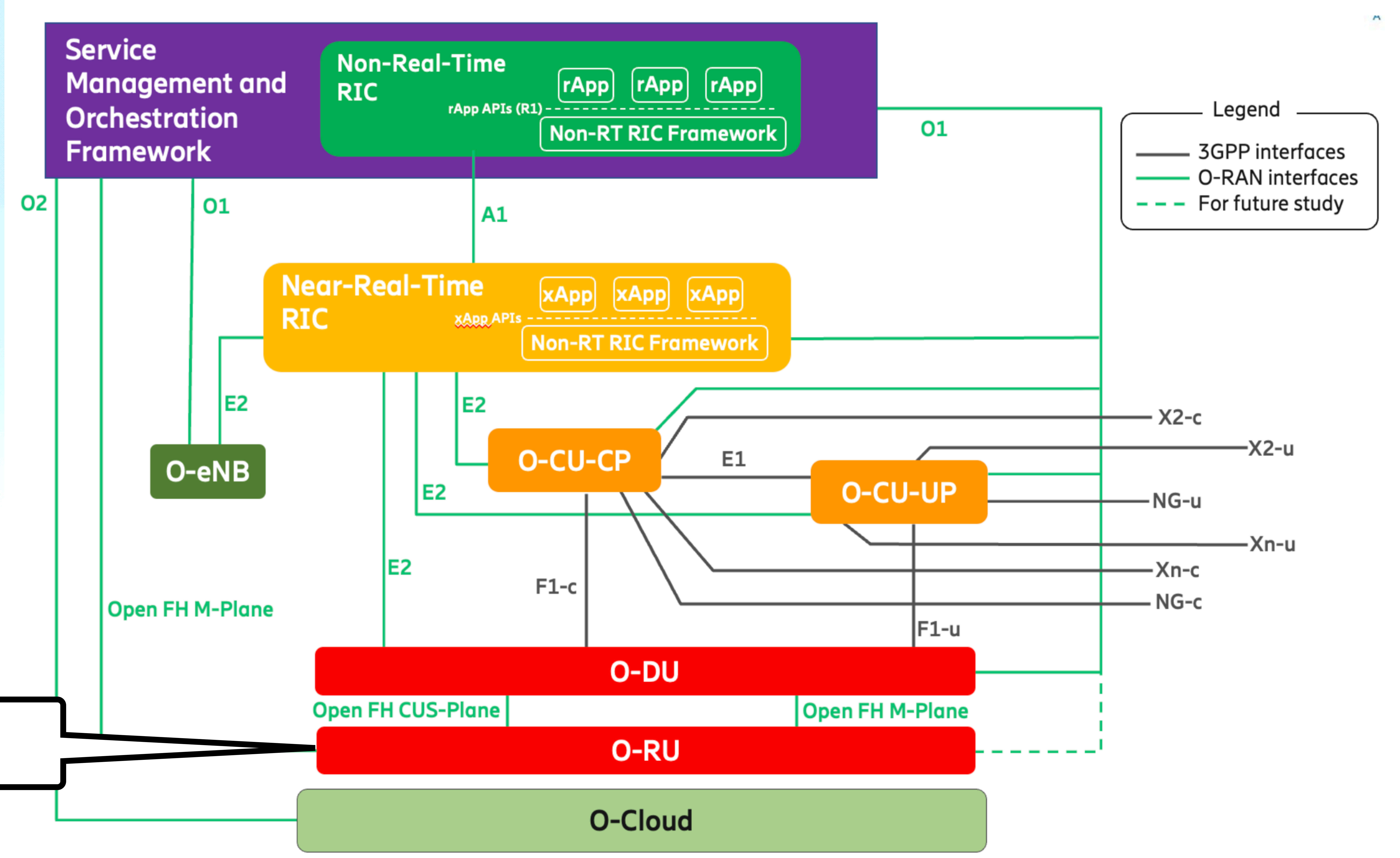
O-RAN Logical Architecture

Cloud Computing platform comprising:

- PHY infra node to host ORAN functions (r.g., RIC, DU, CU)
- Support software components for deployment (OS, VM, container runtime etc.)
- MANO functionality

[O-RAN Architecture Description 12.0]

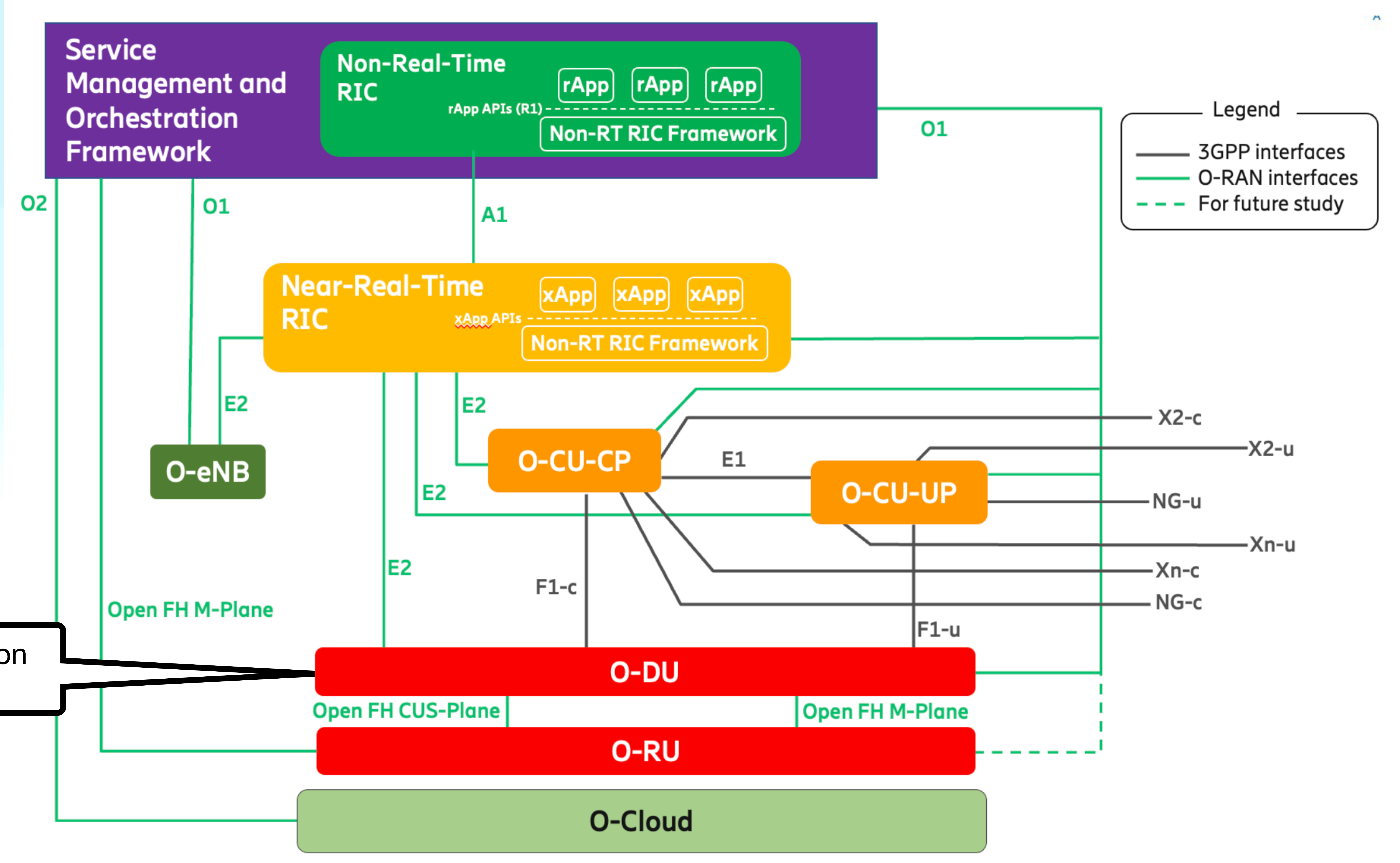
Open RAN Architecture



O-RAN Logical Architecture

Logical node hosting low PHY layer (FFT/iFFT, PRACH, extraction)

Open RAN Architecture

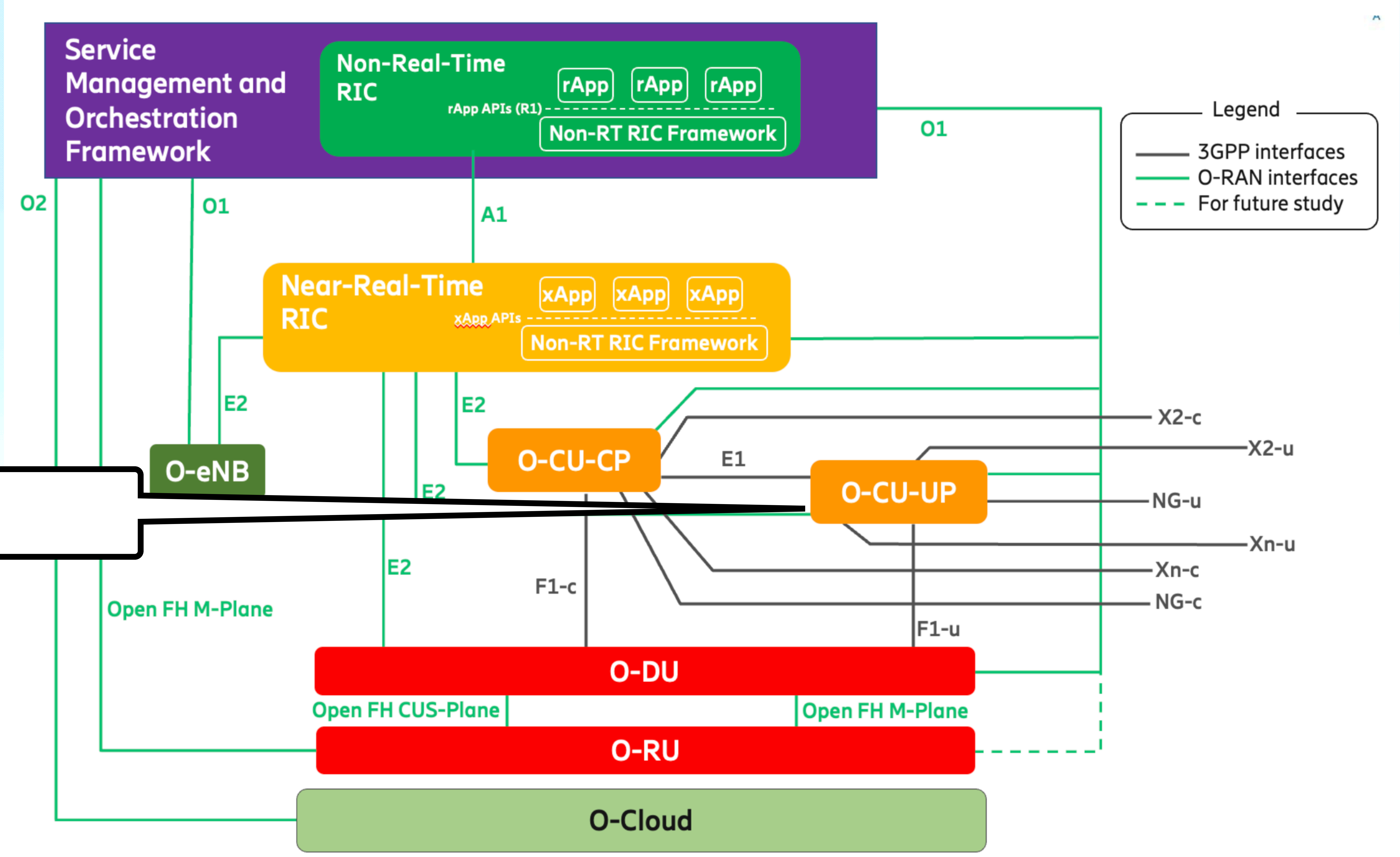


O-RAN Logical Architecture

Logical node hosting RLC/MAC/High-PHY layer based on LLS

Open RAN Architecture

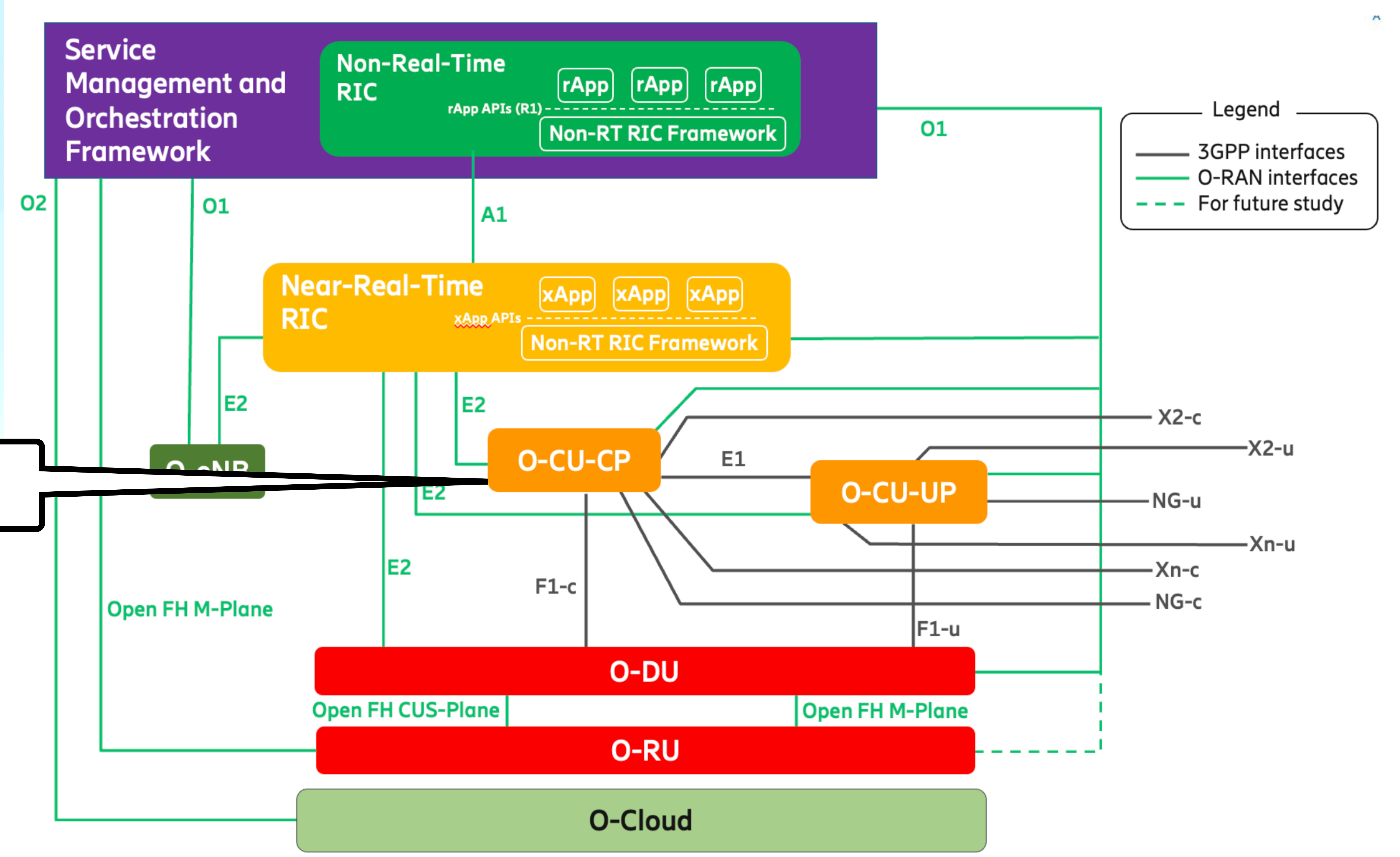
Logical node hosting RRC and UP part of PDCP



O-RAN Logical Architecture

Open RAN Architecture

Logical node hosting RRC and CP part of PDCP

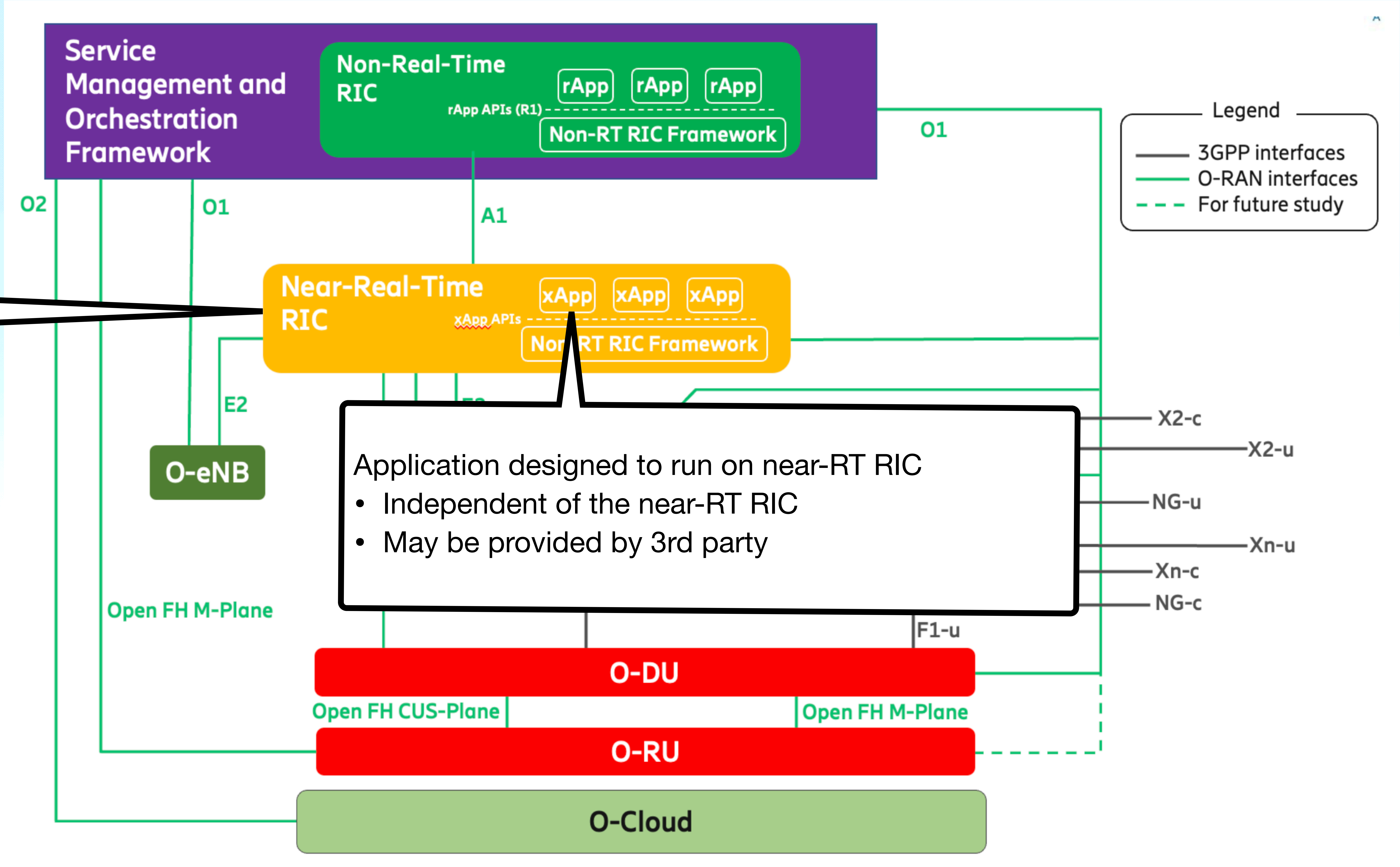


O-RAN Logical Architecture

Open RAN Architecture

Logical node enabling near-RT control/optimisation of RAN elements and resources via fine grained data collection and actions over E2 interface. May include AI/ML workflow.

Application designed to run on near-RT RIC
- Independent of the near-RT RIC
- May be provided by 3rd party



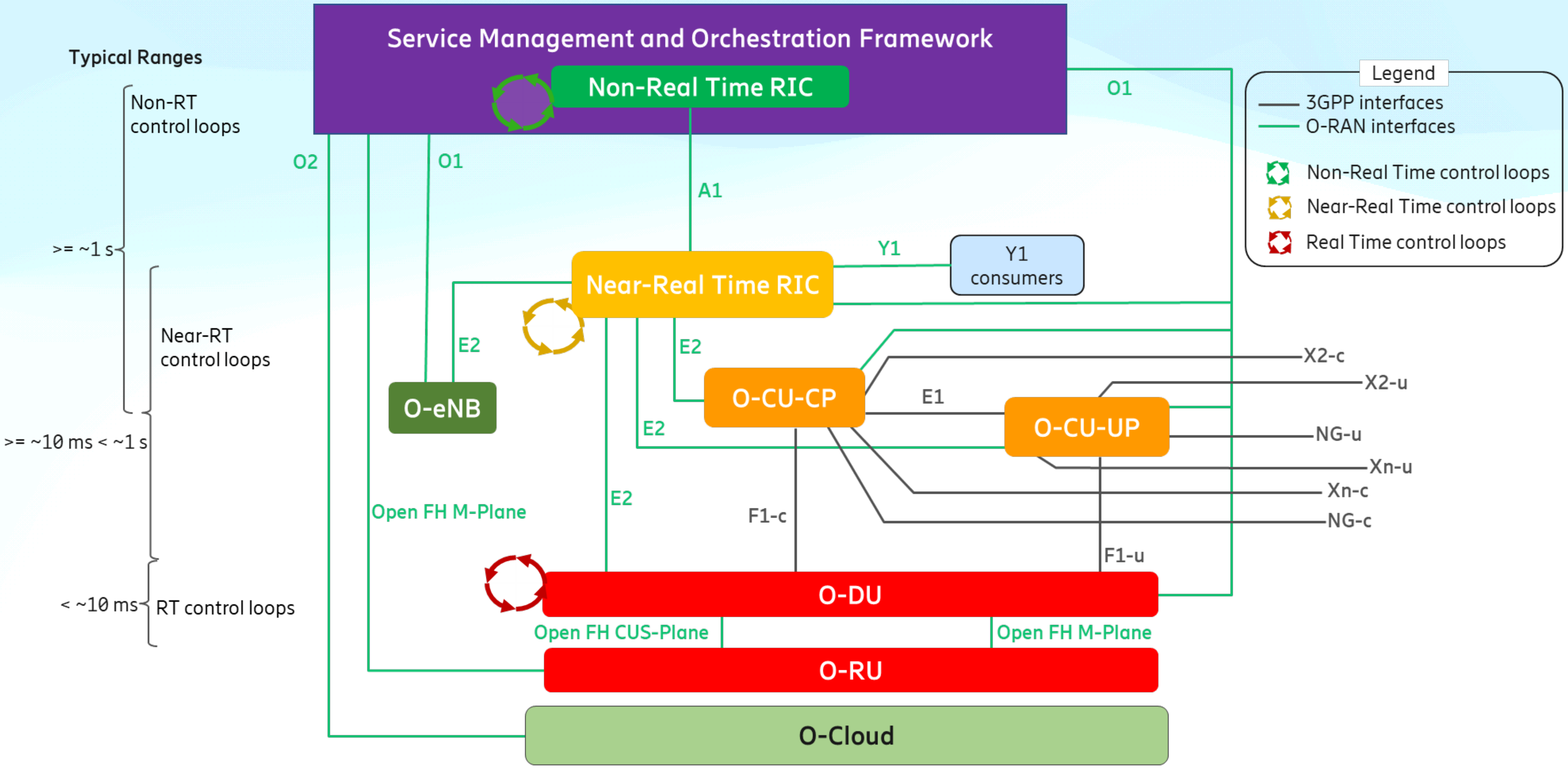
O-RAN Logical Architecture

Control loop

- Service and policy mgmt
- RAN analytics
- AI/ML model training

- RAN Control and optimization
- xApps for use cases
- UE and cell specific metrics

- Real time actions
- Resource management
- Radio scheduling, HARQ,



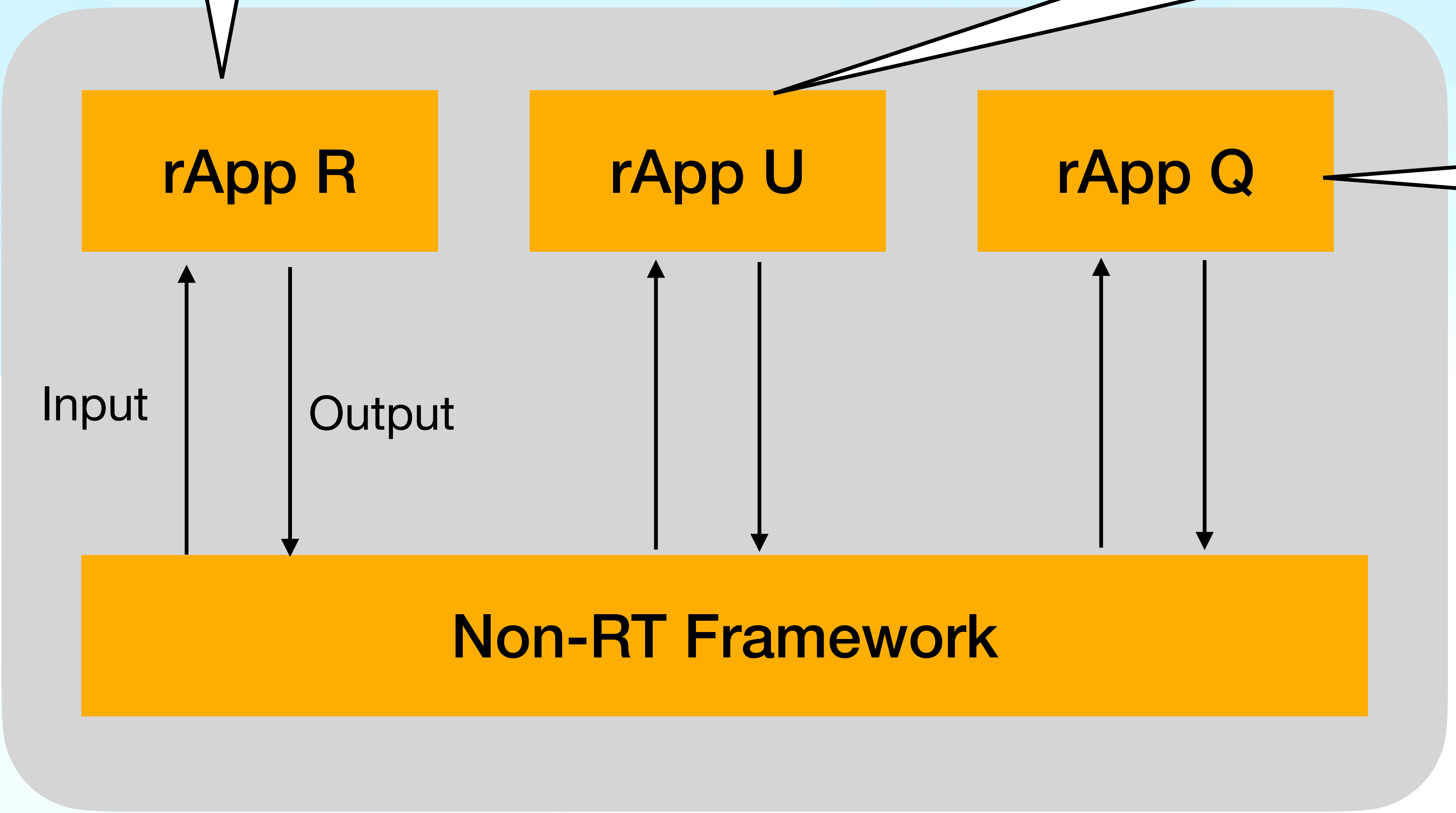
Non-RT RIC example:

Cell Utilisation prediction

- **Input**: Cell utilisation measurement regarding actual capacity utilisation over time for a cell site over time
- **Output**: prediction of cell site utilisation

RF signal Prediction

- **Input**: RF signal experienced by UE for serving or neighbour cell
- **Output**: Prediction of location of UE, Prediction of RF signal



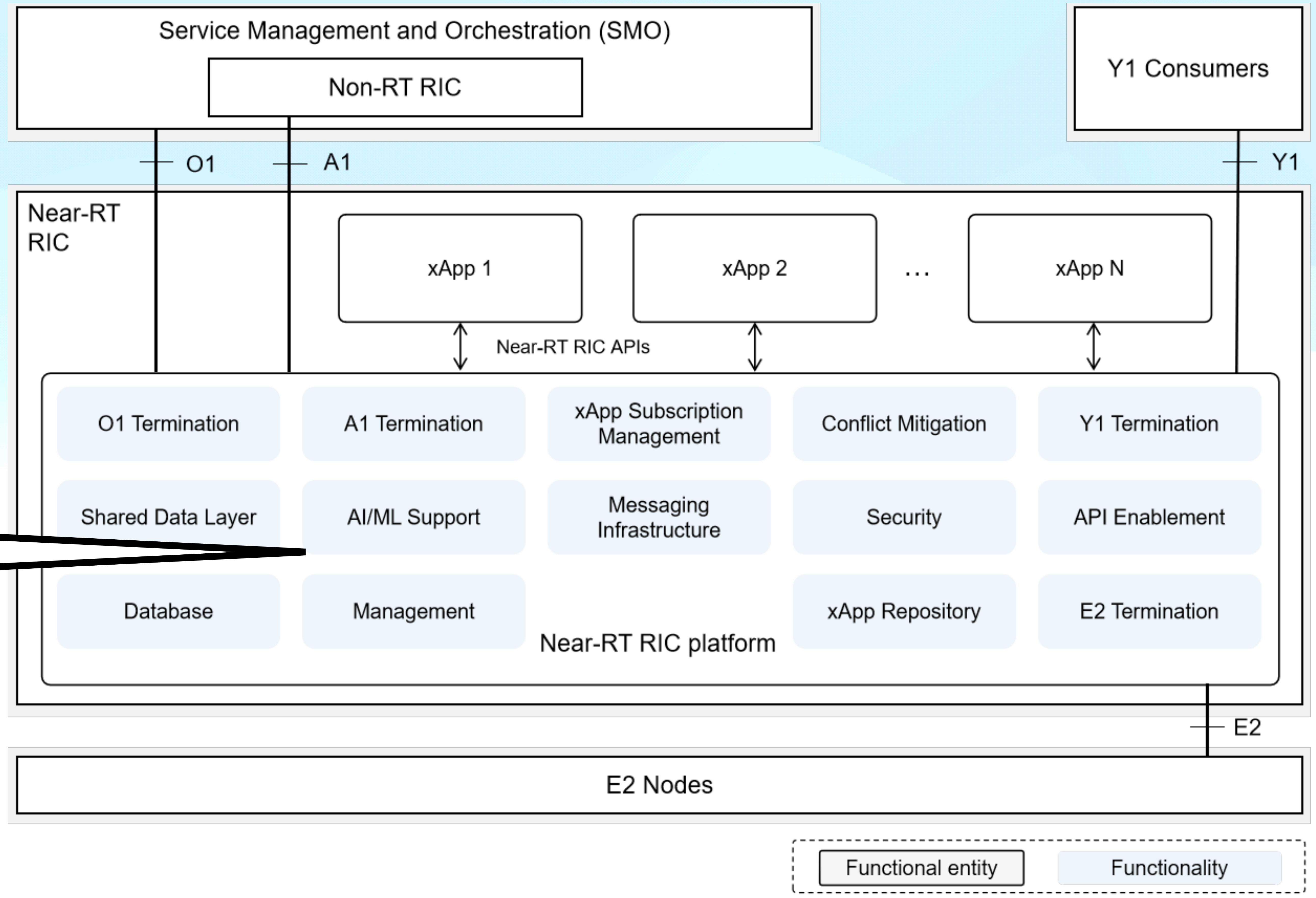
UE QoE prediction

- **Input**: measurement on UE RF signal (actual RAN measurement or prediction), measurement of cell capacity utilisation (actual or predicted)
- **Output**: Calculates QoE experience by particular UE:
  - Estimate actual QoE based on the actual RF signal and actual cell utilisation
  - Estimate QoE in the neighbour cell based RF signal relative to neighbour cell and actual neighbour cell utilisation.
  - Estimate future QoE of serving/neighbour cell based on predicted signal and predicted cell utilisation.

Near-RT RIC Architecture Overview

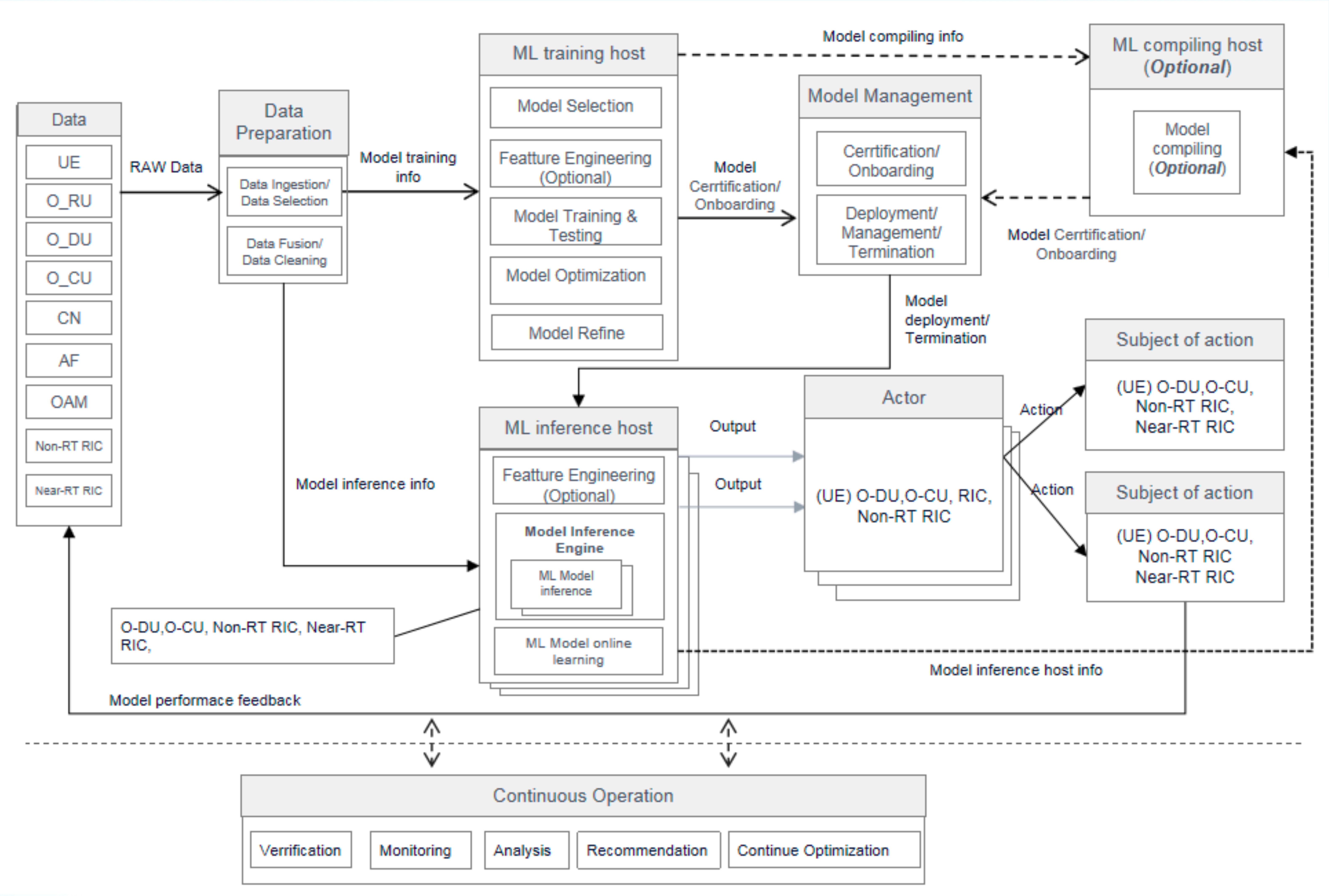
Enables the Near-RT RIC platform with

- **Data pipelining**: enables the Near-RT RIC platform with data ingestion and preparation for xApps.
  - **Input**: E2 node data collected over E2 interface, enrichment information over A1 interface, information from xApps
  - **Output**: The output is data sets that are ready to be consumed by AI/ML Model.
- **Model management:** Offers storage, retrieval, and version control of AI/ML models for xApps
- **Training**: This functionality enables training of AI/ML Models for xApps within Near-RT RIC.
- **Inference:** Near-RT RIC platform offers inference of AI/ML Models for xApps. The associated AI/ML models are managed by the Near-RT RIC platform



[O-RAN AI/ML workflow description and requirements]

Use of the ML components and terminologies



[O-RAN AI/ML workflow description and requirements]

ML Model Lifecycle Implementation Example

1. ML Modeller uses a designer environment along with ML toolkits (e.g.,TensorFlow) to create the initial ML model

2. The initial model is sent to training hosts for training

3. The appropriate data sets are collected from the Near-RT RIC, O-CU and O-DU to a data lake and passed to the ML training hosts.

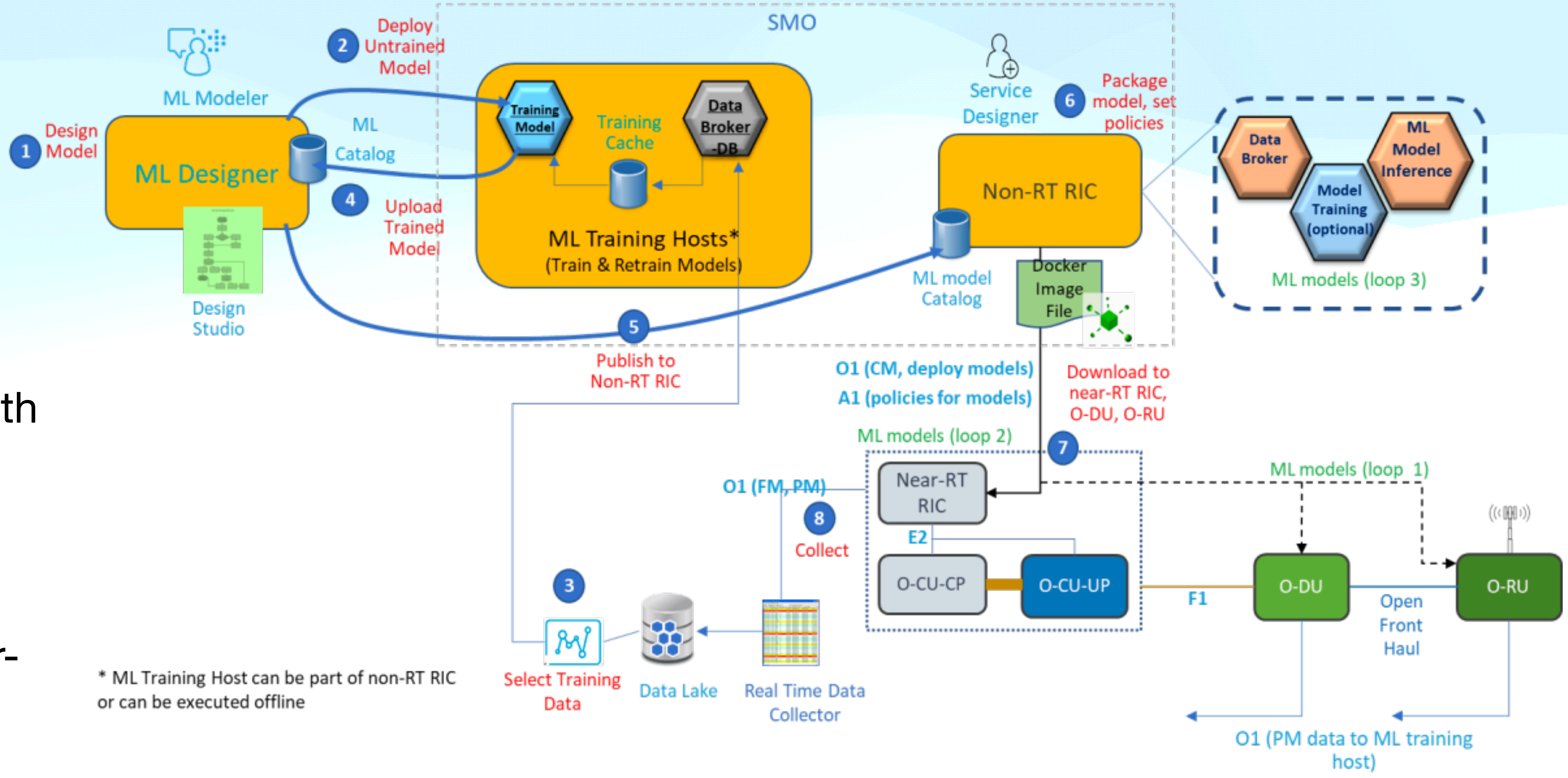
4. The trained model/sub models are uploaded to the ML designer catalog

5. The ML model is published to Non-RT RIC along with the the associated license and metadata

6. Non-RT RIC creates a containerized ML application

7. Non-RT RIC deploys the ML application to the Near-RT RIC, O-DU and O-CU using the O1 interface.

8. PM data is sent back to ML training hosts from Near-RT RIC, O-DU and O-CU for retraining



* ML Training Host can be part of non-RT RIC or can be executed offline

AI/ML Models in O-RAN Use Cases

Use Case	AI/ML models functionality description	AI/ML algorithms types (example)	Data Input	Data Output
QoE Optimization	Service type classification (eMBB, URLLC, mMTC)	Supervised learning (e.g., CNN, DNN)	User traffic data	service type
	KQI/QoE prediction (e.g., good, bad or video stall ratio, duration)	Supervised learning (e.g, LSTM)	**Network data:** • L2 measurement report related to traffic pattern, e.g., throughput, latency, packets per second • UE level radio channel information, mobility related metrics • RAN protocol stack status: e.g. PDCP buffer status • Cell level information: e.g. DL/UL PRB occupation rate **Application data:** • Video QoE score • Video initial delay • Stalling detail including the timestamp stalling duration, stalling ratio	KQI/QoE value e.g., good/bad, stalling ratio, video stalling duration, vMoS value
	Available radio bandwidth prediction	Supervised learning (e.g., DNN)	Similar to above	Available radio Bandwidth
Traffic Steering	A cell load prediction/user traffic volume prediction	Supervised learning (time series prediction, e.g., SVR, DNN)	load related counters, e.g., UL/DL PRB occupation	Same as input
	Radio finger print prediction	A supervised learning (e.g., SVR, GBDT)	A Intra-frequency MR data and PM counters, e.g., RSRP, RSRQ, MCS, CQI, etc	Same as input

O-RAN AI/ML workflow description and requirements

Challenges for ORAN

System integration	Security Risks	Widespread adoption	Questionable CAPEX and OPEX saving
• May lead towards some degree of vendor locking again.	• Zero trust principles guiding O-RAN security requirements.	• 35% of UK traffic through ORAN by 2030	• May save initial CAPEX and OPEX. • Integration may be costly in the long run.

DCMS-UK ORAN traffic ambition for 2030

Challenges for AI in RAN

Data management

- Managing a huge amount of data with efficiency and security, especially user data.

AI tools for lifecycle management

- CSPs need to consider future-proof tools in the existing deployments.

Deploying new AI capabilities in their systems

- These are retraining, assurance, explainability, and experimentation, which require optimized AI/ML

Optimal hardware considerations

- There are new hardware requirements to satisfy AI implementation including storage capacity and processing power.
- Selecting platforms that effectively support these

Thank you