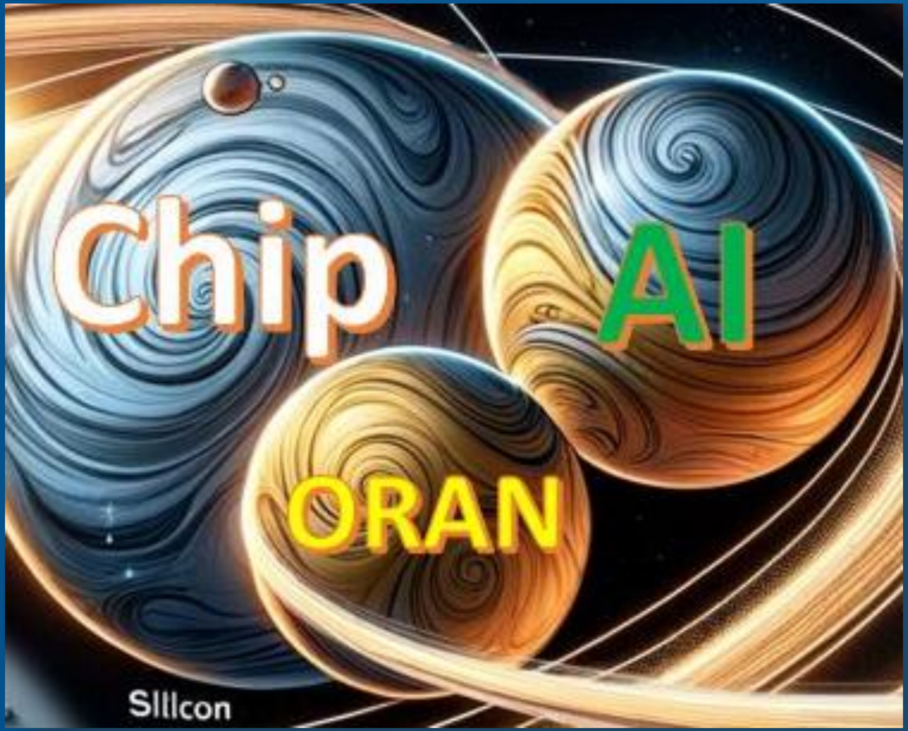
O-RAN
ALLIANCE

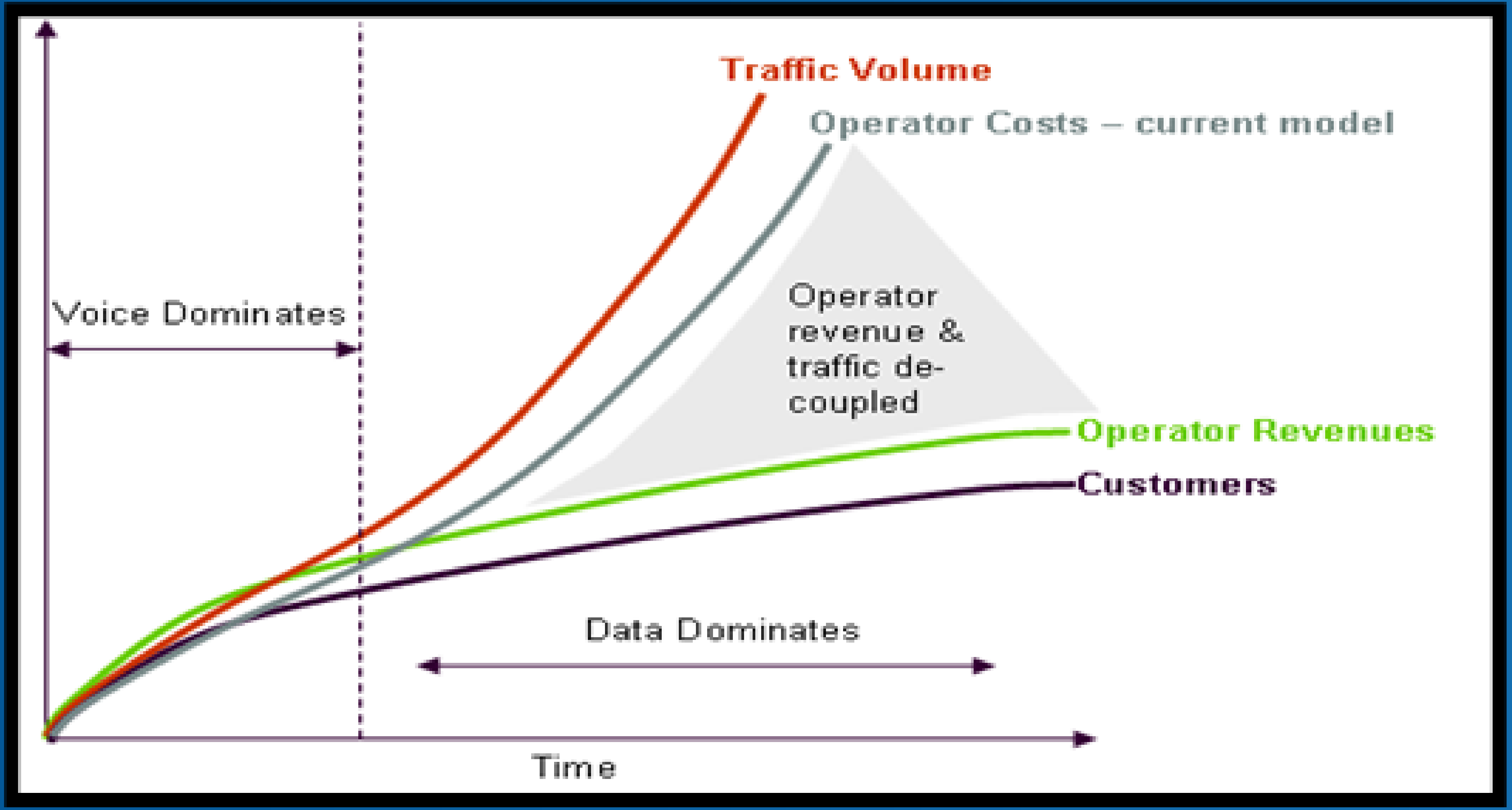


# Confluence of ORAN, AI and Chip

*Alex Jinsung Choi*

Chair of O-RAN ALLIANCE
SVP Deutsche Telekom

# Discrepancies between traffic growth and revenue growth (Source: Accenture)



https://telecom.altanai.com/category/access-and-physical-layer/

# Applying Cost Criteria when prioritizing work items

New feature addition
Capacity increase/QoS/Latency improving
Techs requiring additional resources and complex control mechanisms, spectrum-RAT dedication, High dimensional signal processing  etc.

Cost Accelerators

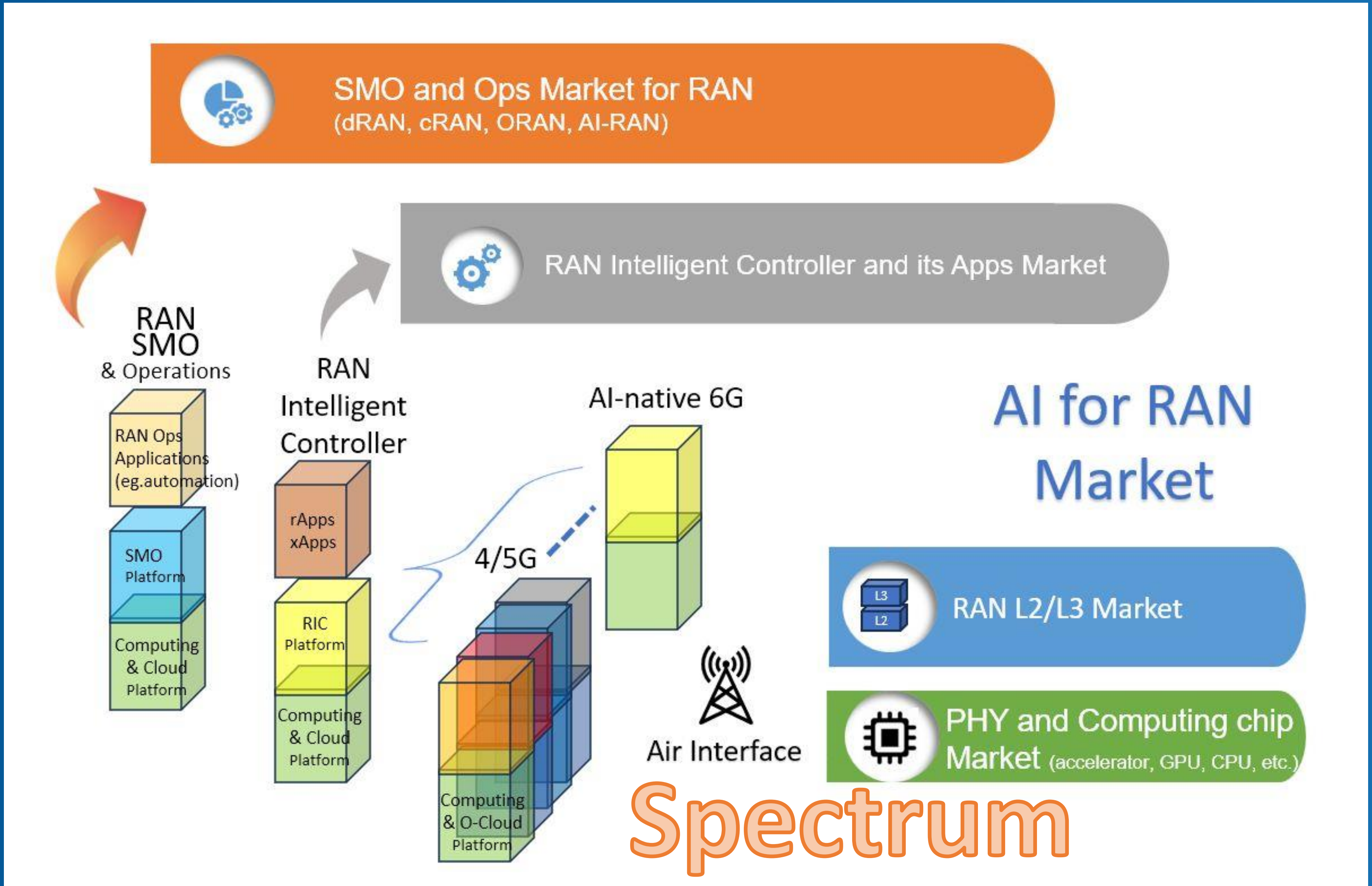
Techs for Cost Optimization drivers: commoditization, NG OSS, merchant silicons, simplification, resource sharing, OPEX driven, automation
AIOps, DevOps, etc.

Cost Optimizers

**Cost**

Spectrum Cost
RAN Cost
xHAUL Cost
Cloud Infra Cost
SW Licensing Cost
Site Cost
Energy Cost
Operational Cost
Integration Cost
AI/ML infra cost
etc.

# O-RAN Landscape



SMO and Ops Market for RAN
(dRAN, cRAN, ORAN, AI-RAN)

RAN Intelligent Controller and its Apps Market

AI for RAN Market

RAN L2/L3 Market

PHY and Computing chip Market (accelerator, GPU, CPU, etc.)

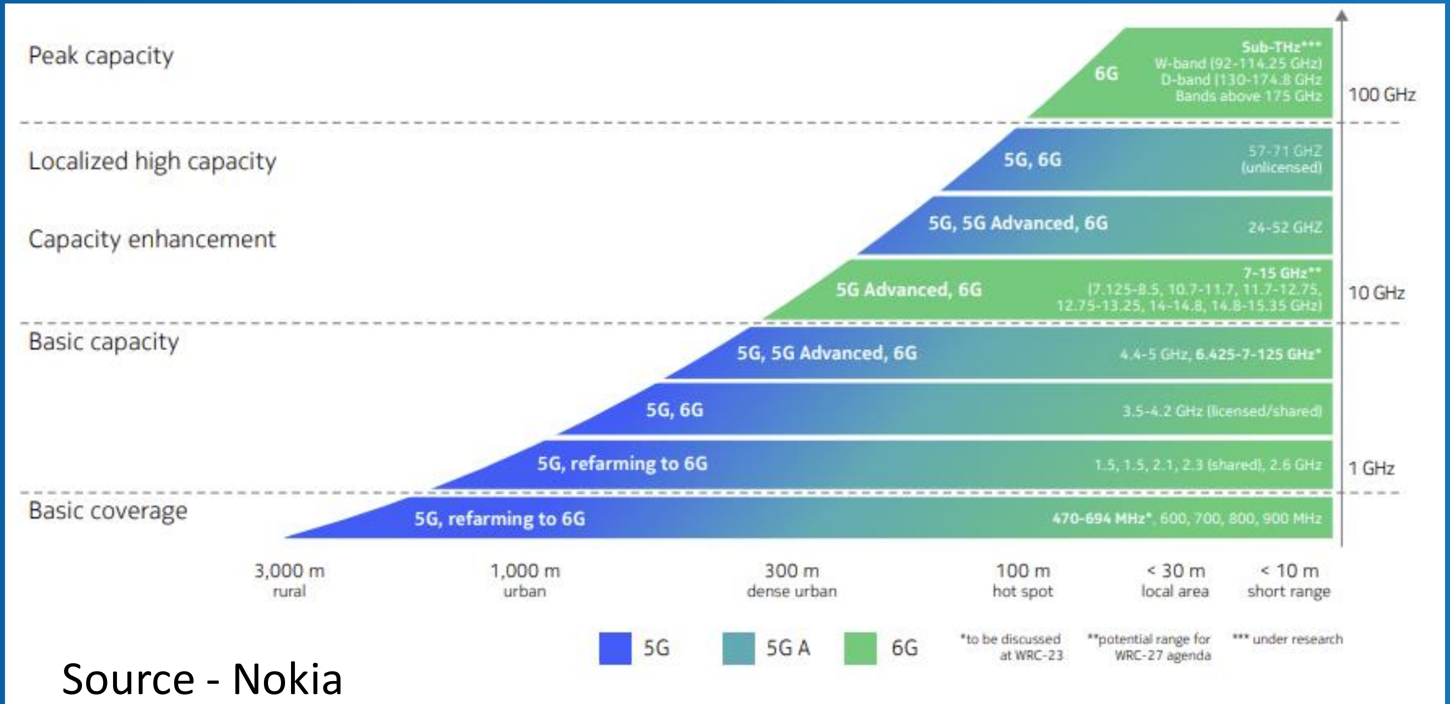
Spectrum

RAN SMO
& Operations
RAN Ops Applications (eg.automation)
SMO Platform
Computing & Cloud Platform
RAN Intelligent Controller
rApps xApps
RIC Platform
Computing & Cloud Platform
AI-native 6G
4/5G
Computing & O-Cloud Platform
Air Interface
L3
L2

# RAN - Spectrum Engineering

Spectrum engineering is defined as the discipline within telecommunications focused on the strategic management and technical optimization of radio frequency (RF) spectrum. It involves the analysis, planning, and regulation of the spectrum to ensure efficient, fair, and effective use of this finite resource. Key activities in spectrum engineering include:

- Spectrum Modelling and Analysis
- Interference Management
- Spectrum Sharing and Re-farming
- Coordination and Compliance
- Technical Analysis and Simulation

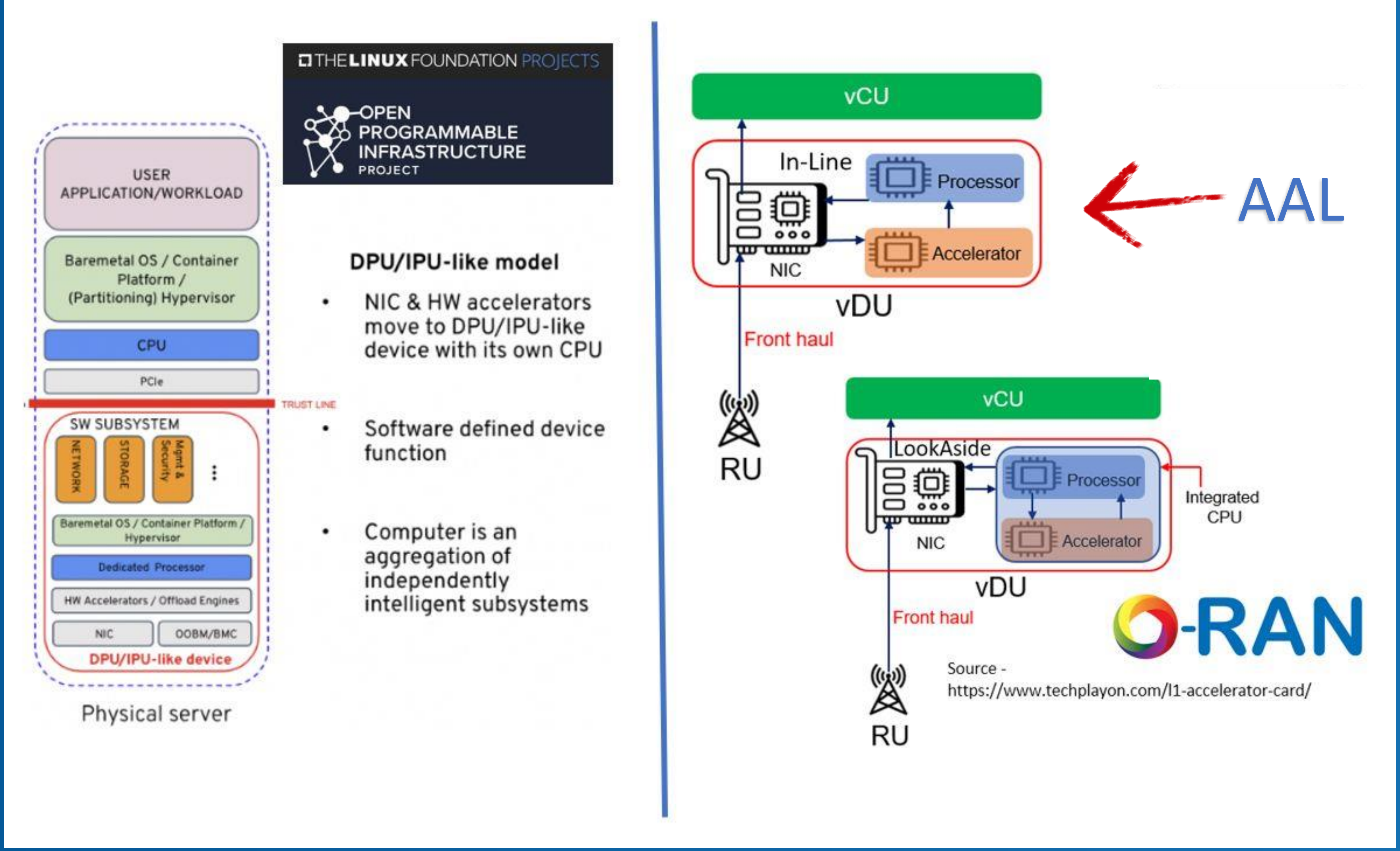


Source - Nokia

- Support for Emerging Technologies: As new technologies such as Dynamic Spectrum Sharing, MRSS, Advanced Cognitive Radio/Software Defined Radio, Sensing in 5G/5G-Advanced/6G

# O-RAN Accelerator Adaption Layer and Silicon

## why relevant?

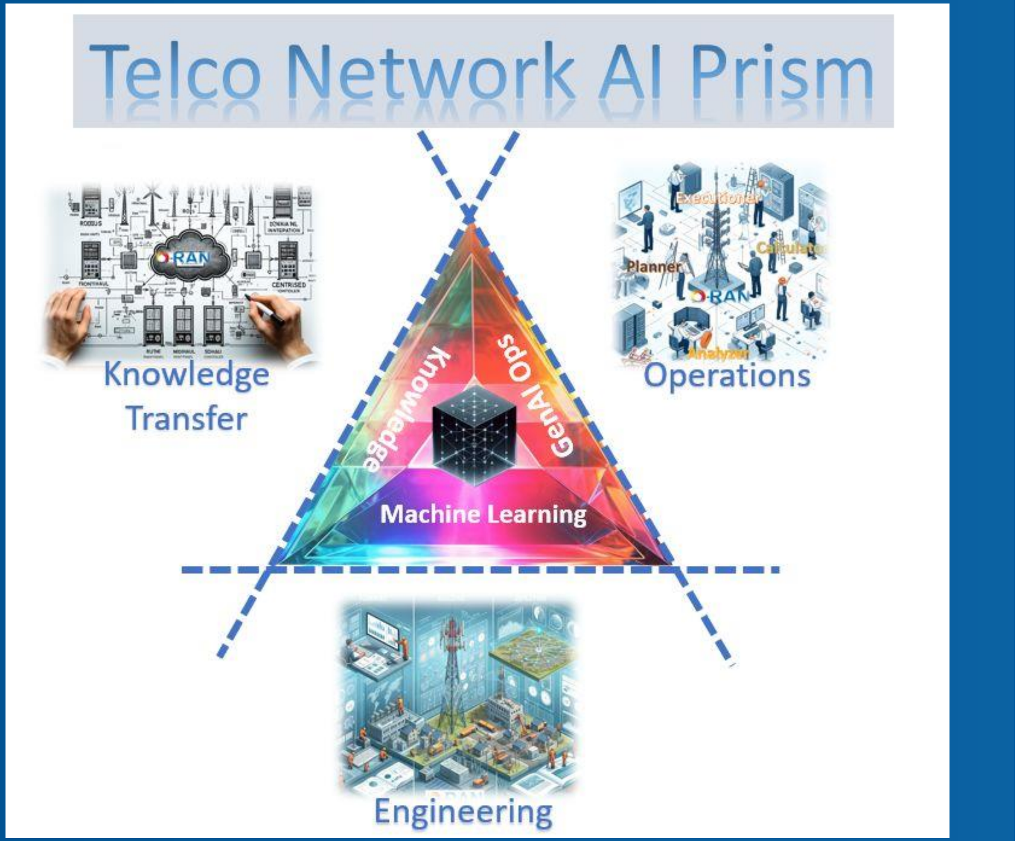


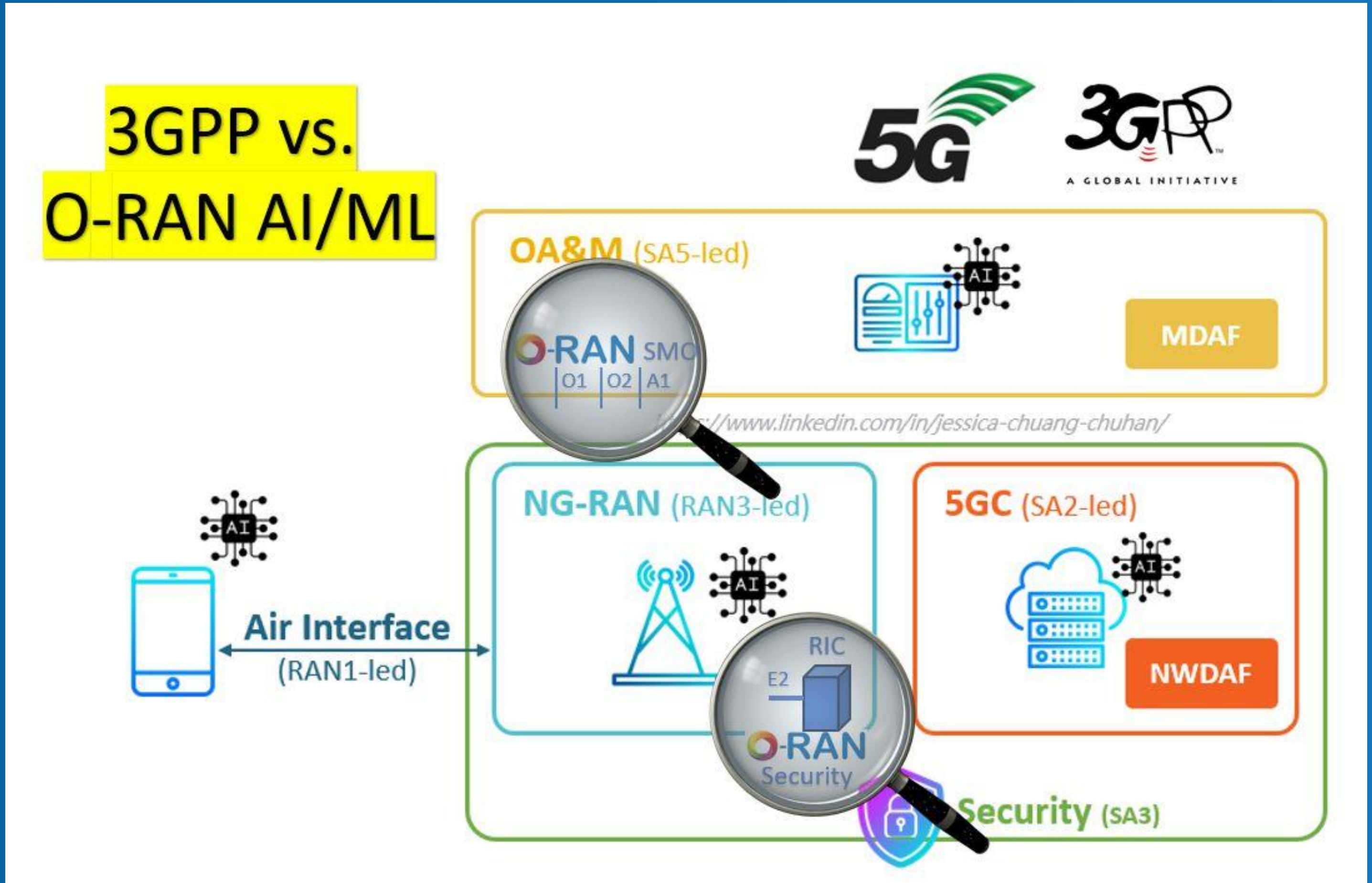
**DPU/IPU-like model**

- NIC & HW accelerators move to DPU/IPU-like device with its own CPU
- Software defined device function
- Computer is an aggregation of independently intelligent subsystems

Source - https://www.techplayon.com/l1-accelerator-card/

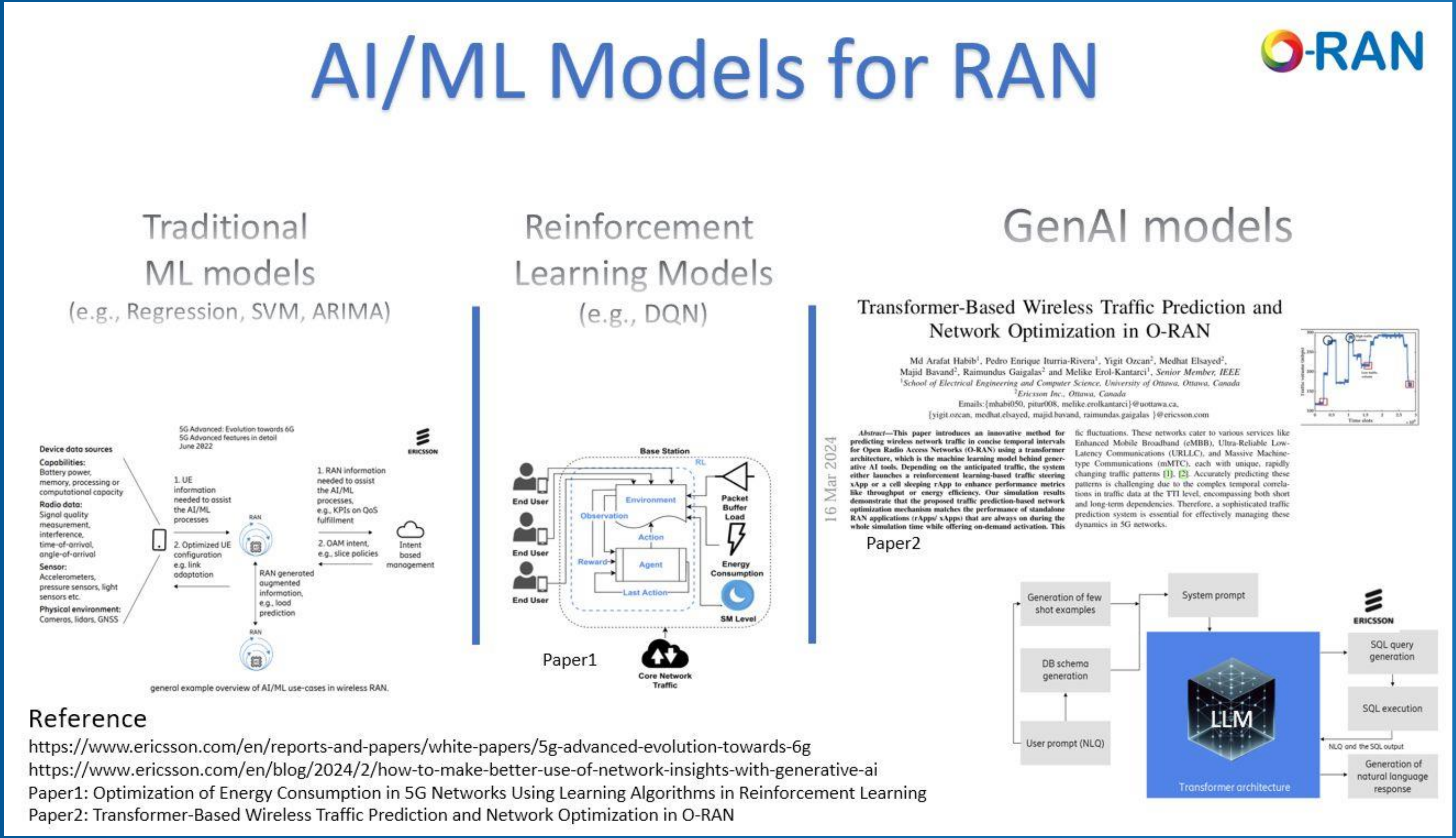
# AI for RAN





Note - **Illustration created by adding O-RAN AI/ML to the picture in Jessica Chuang's great LinkedIn post with the title of 'AI in 5G Use Case - Summary of 3GPP's Work & Study on AI/ML for 5G System'**

# Three Types of AI/ML Models for RAN



## AI/ML Models for RAN

### Traditional ML models

(e.g., Regression, SVM, ARIMA)

5G Advanced: Evolution towards 6G
5G Advanced features in detail
June 2022
ERICSSON
Device data sources
Capabilities:
Battery power, memory, processing or computational capacity
Radio data:
Signal quality measurement, interference, time-of-arrival, angle-of-arrival
Sensor:
Accelerometers, pressure sensors, light sensors etc.
Physical environment:
Cameras, lidars, GNSS
1. UE information needed to assist the AI/ML processes
2. Optimized UE configuration e.g. link adaptation
RAN
1. RAN information needed to assist the AI/ML processes, e.g., KPIs on QoS fulfillment
2. OAM intent, e.g., slice policies
Intent based management
RAN generated augmented information, e.g., load prediction
RAN

general example overview of AI/ML use-cases in wireless RAN.

### Reinforcement Learning Models

(e.g., DQN)

Base Station
RL
End User
End User
End User
Environment
Observation
Action
Reward
Agent
Last Action
Packet Buffer Load
Energy Consumption
SM Level
Core Network Traffic

Paper1

### GenAI models

Transformer-Based Wireless Traffic Prediction and Network Optimization in O-RAN

Md Arafat Habib[1], Pedro Enrique Iturria-Rivera[1], Yigit Ozcan[2], Medhat Elsayed[2], Majid Bavand[2], Raimundas Gaigalas[2] and Melike Erol-Kantarci[1], *Senior Member, IEEE*
[1]*School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, Canada*
[2]*Ericsson Inc., Ottawa, Canada*
Emails:{mhabi050, pitur008, melike.erolkantarci}@uottawa.ca,
{yigit.ozcan, medhat.elsayed, majid.bavand, raimundas.gaigalas }@ericsson.com

16 Mar 2024

***Abstract*—This paper introduces an innovative method for predicting wireless network traffic in concise temporal intervals for Open Radio Access Networks (O-RAN) using a transformer architecture, which is the machine learning model behind generative AI tools. Depending on the anticipated traffic, the system either launches a reinforcement learning-based traffic steering xApp or a cell sleeping rApp to enhance performance metrics like throughput or energy efficiency. Our simulation results demonstrate that the proposed traffic prediction-based network optimization mechanism matches the performance of standalone RAN applications (rApps/ xApps) that are always on during the whole simulation time while offering on-demand activation. This**

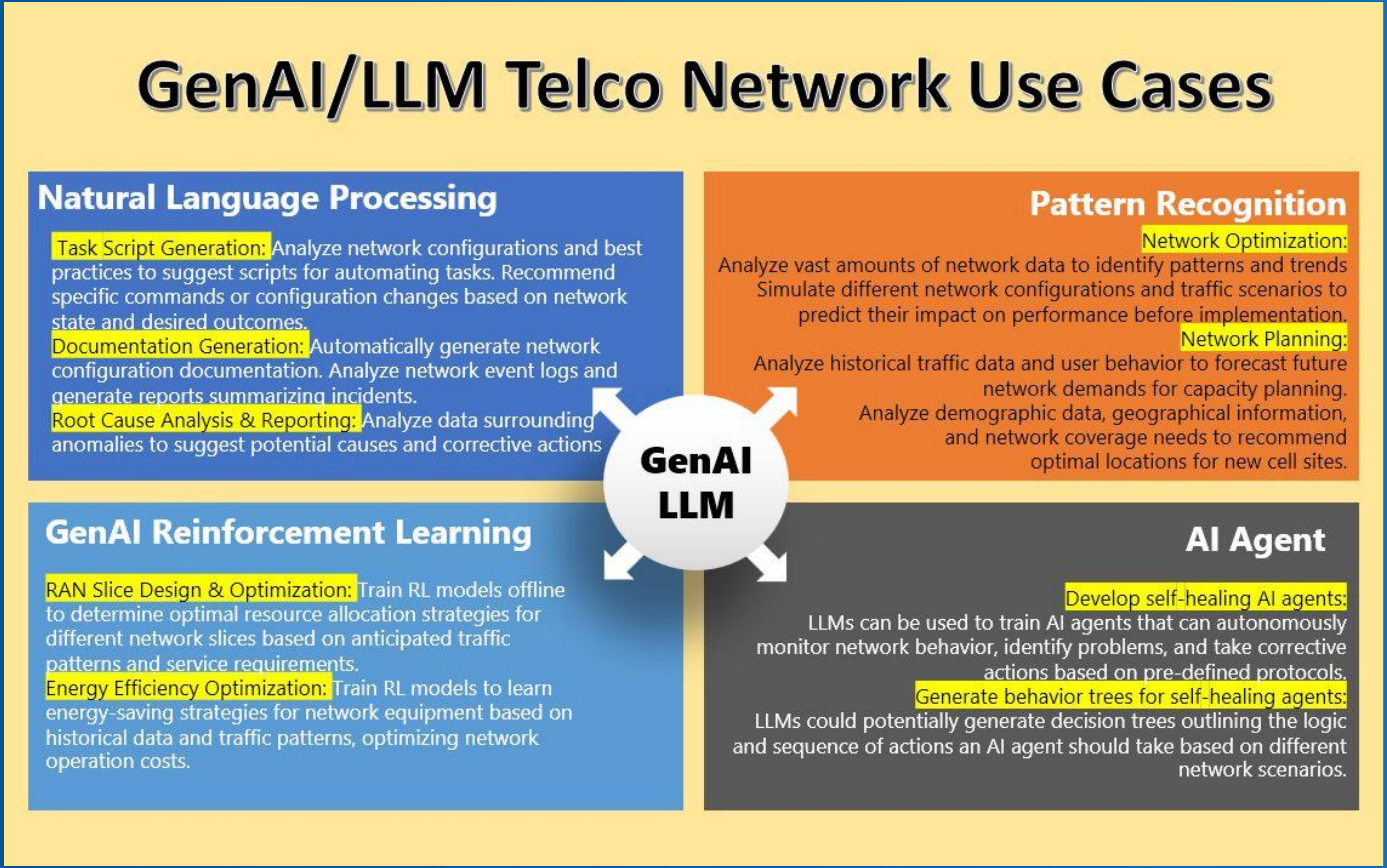
fic fluctuations. These networks cater to various services like Enhanced Mobile Broadband (eMBB), Ultra-Reliable Low-Latency Communications (URLLC), and Massive Machine-type Communications (mMTC), each with unique, rapidly changing traffic patterns [1], [2]. Accurately predicting these patterns is challenging due to the complex temporal correlations in traffic data at the TTI level, encompassing both short and long-term dependencies. Therefore, a sophisticated traffic prediction system is essential for effectively managing these dynamics in 5G networks.

Paper2

Reference

https://www.ericsson.com/en/reports-and-papers/white-papers/5g-advanced-evolution-towards-6g
https://www.ericsson.com/en/blog/2024/2/how-to-make-better-use-of-network-insights-with-generative-ai
Paper1: Optimization of Energy Consumption in 5G Networks Using Learning Algorithms in Reinforcement Learning
Paper2: Transformer-Based Wireless Traffic Prediction and Network Optimization in O-RAN

# AI for RAN use cases including RAN-Ops



## GenAI/LLM Telco Network Use Cases

### Natural Language Processing

Task Script Generation: Analyze network configurations and best practices to suggest scripts for automating tasks. Recommend specific commands or configuration changes based on network state and desired outcomes.

Documentation Generation: Automatically generate network configuration documentation. Analyze network event logs and generate reports summarizing incidents.

Root Cause Analysis & Reporting: Analyze data surrounding anomalies to suggest potential causes and corrective actions

### Pattern Recognition

Network Optimization:

Analyze vast amounts of network data to identify patterns and trends Simulate different network configurations and traffic scenarios to predict their impact on performance before implementation.

Network Planning:

Analyze historical traffic data and user behavior to forecast future network demands for capacity planning.

Analyze demographic data, geographical information, and network coverage needs to recommend optimal locations for new cell sites.

GenAI LLM

### GenAI Reinforcement Learning

RAN Slice Design & Optimization: Train RL models offline to determine optimal resource allocation strategies for different network slices based on anticipated traffic patterns and service requirements.

Energy Efficiency Optimization: Train RL models to learn energy-saving strategies for network equipment based on historical data and traffic patterns, optimizing network operation costs.

### AI Agent

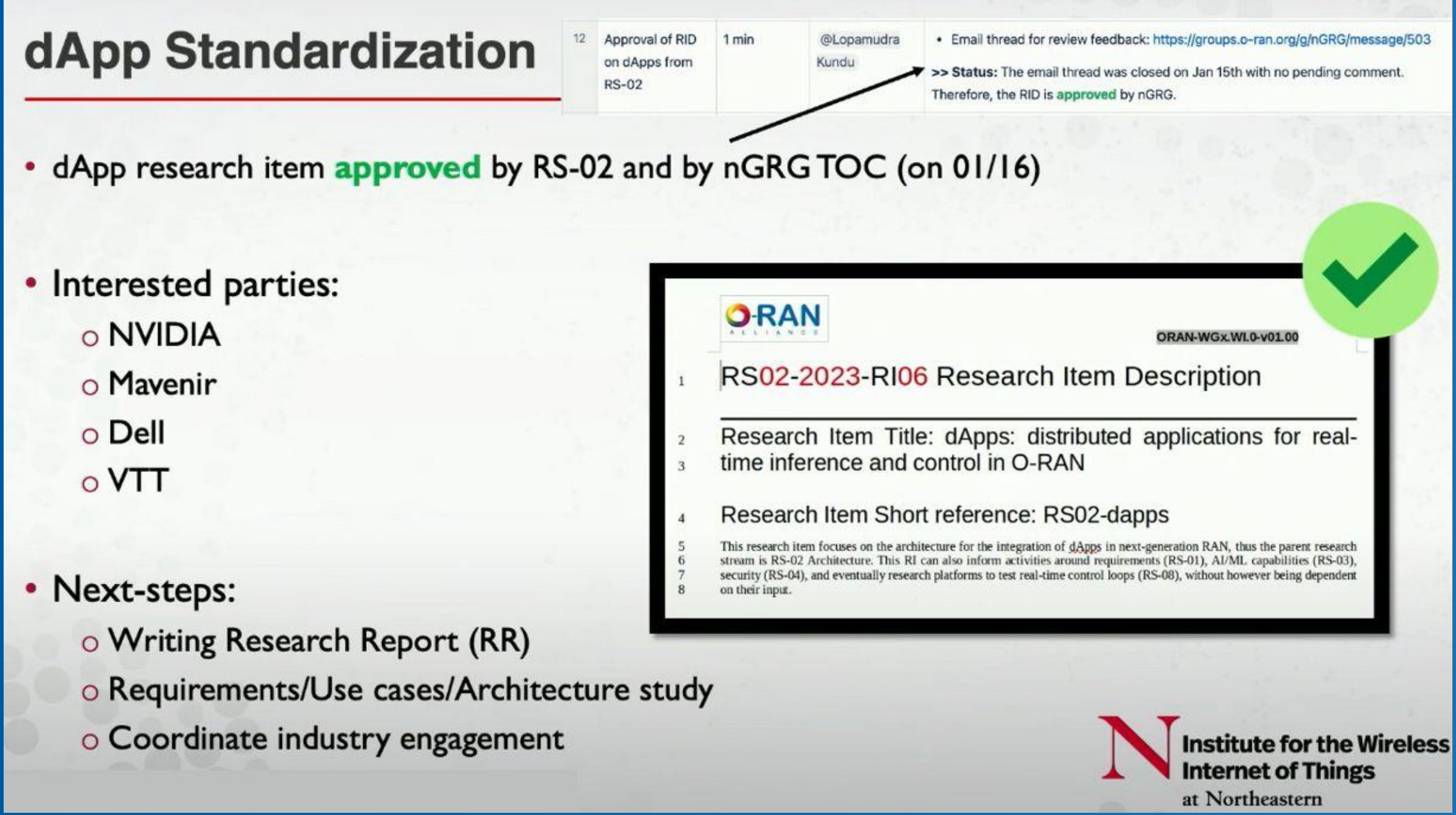
Develop self-healing AI agents:

LLMs can be used to train AI agents that can autonomously monitor network behavior, identify problems, and take corrective actions based on pre-defined protocols.

Generate behavior trees for self-healing agents:

LLMs could potentially generate decision trees outlining the logic and sequence of actions an AI agent should take based on different network scenarios.

# Advanced RIC Apps - Real Time RAN Control



## dApp Standardization

| | | | | |
|---|---|---|---|---|
| 12 | Approval of RID on dApps from RS-02 | 1 min | @Lopamudra Kundu | • Email thread for review feedback: https://groups.o-ran.org/g/nGRG/message/503<br>>> **Status:** The email thread was closed on Jan 15th with no pending comment. Therefore, the RID is approved by nGRG. |

- dApp research item **approved** by RS-02 and by nGRG TOC (on 01/16)

- Interested parties:
  - NVIDIA
  - Mavenir
  - Dell
  - VTT

- Next-steps:
  - Writing Research Report (RR)
  - Requirements/Use cases/Architecture study
  - Coordinate industry engagement

ORAN-WGx.WI.0-v01.00

RS02-2023-RI06 Research Item Description

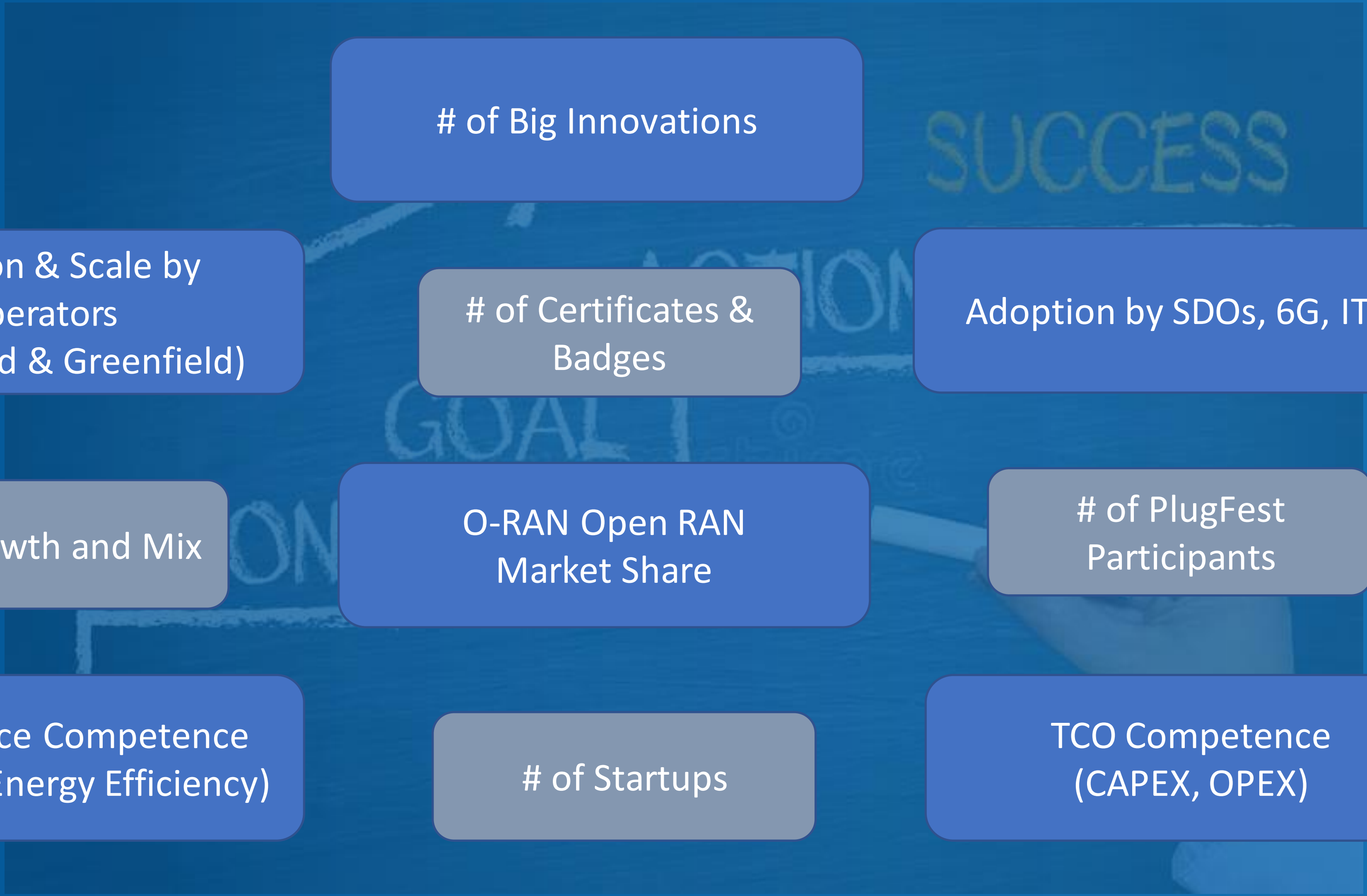
Research Item Title: dApps: distributed applications for real-time inference and control in O-RAN

Research Item Short reference: RS02-dapps

This research item focuses on the architecture for the integration of dApps in next-generation RAN, thus the parent research stream is RS-02 Architecture. This RI can also inform activities around requirements (RS-01), AI/ML capabilities (RS-03), security (RS-04), and eventually research platforms to test real-time control loops (RS-08), without however being dependent on their input.

Institute for the Wireless Internet of Things at Northeastern

# How to measure success of O-RAN Alliance



# of Big Innovations

Adoption & Scale by Operators (Brownfield & Greenfield)

# of Certificates & Badges

Adoption by SDOs, 6G, ITU

Org Growth and Mix

O-RAN Open RAN Market Share

# of PlugFest Participants

Performance Competence (incl. Cost- Energy Efficiency)

# of Startups

TCO Competence (CAPEX, OPEX)

O-RAN

ALLIANCE