



July 20th, 2021
(IEEE 7th WF-IoT)

Application of AI/ML for Learning-based Scaling for the Industry 4.0 Services

*IEEE 7th World Forum on Internet of Things
TOP1: Artificial Intelligence and Machine Learning*

Xi Li
5G/6G R&D, NEC Labs Europe
E-mail: xi.li@neclab.eu



Content

- Introduction
 - Requirements of IoTs in Industries
 - 5Growth Architecture & Innovations
- Closed-Loop Automation Framework
- Use Cases of Applying AI/ML
- AI/ML for Automated Service Scaling
 - Use case and implementation
 - Proof-of-Concept Results
- Key Take-Aways

Requirements of IoTs in Industries

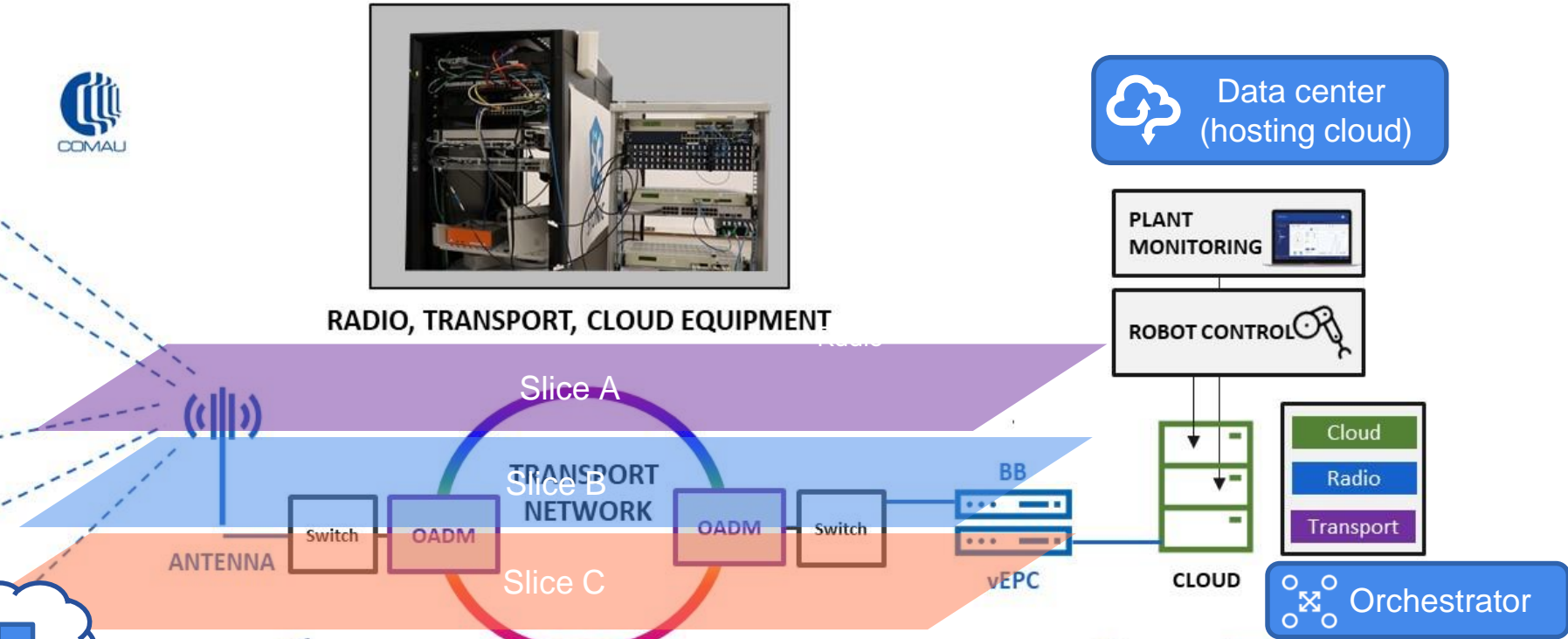
Industrial Applications

Digital Twin
Critical IoT

Telemetry & Monitoring
Massive IoT

Remote Support
Broadband IoT

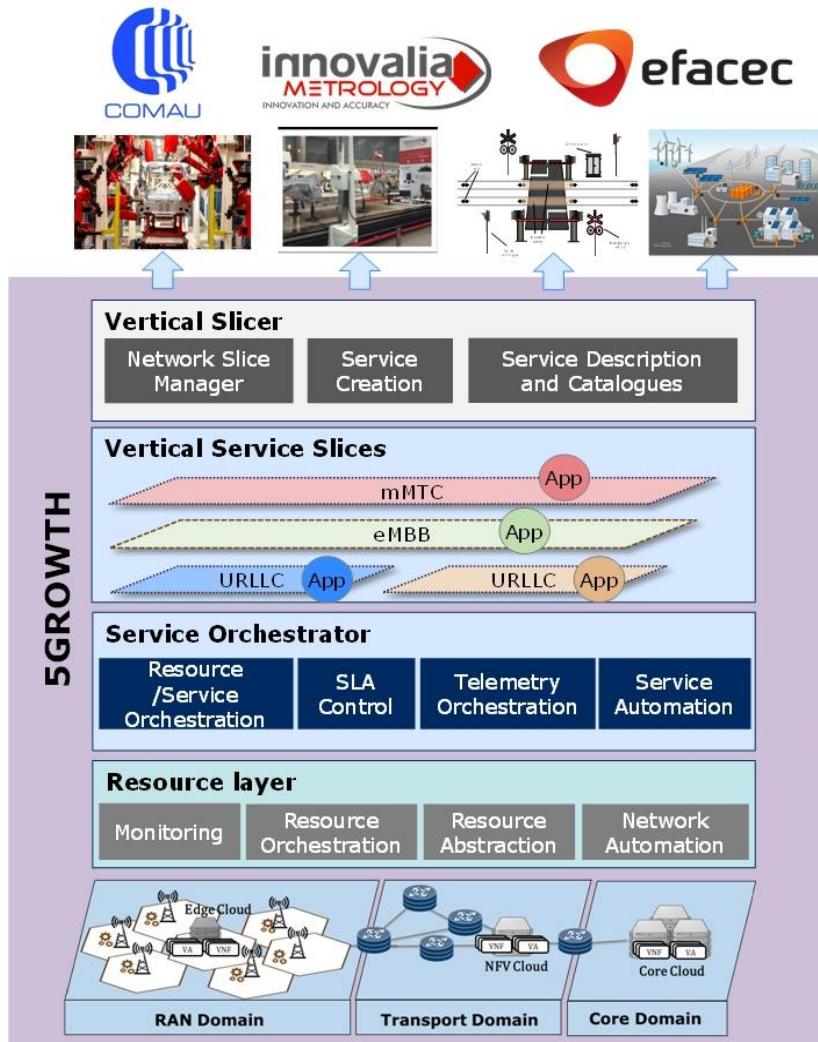
5G End-to-End Infrastructure (Radio, Transport, Cloud)



- In a future smart factory environment, there will be increasing number of IoT devices and also co-existing of different IoT applications, requiring cost-efficient use of the available infrastructure resources and adaptive to the varying service demands over time.
- The dynamics of such systems lead to the need of smart resource allocation throughout the service lifetime to provide a continuous and automated assurance of the service quality to control the IoT applications.

AI-driven Innovations

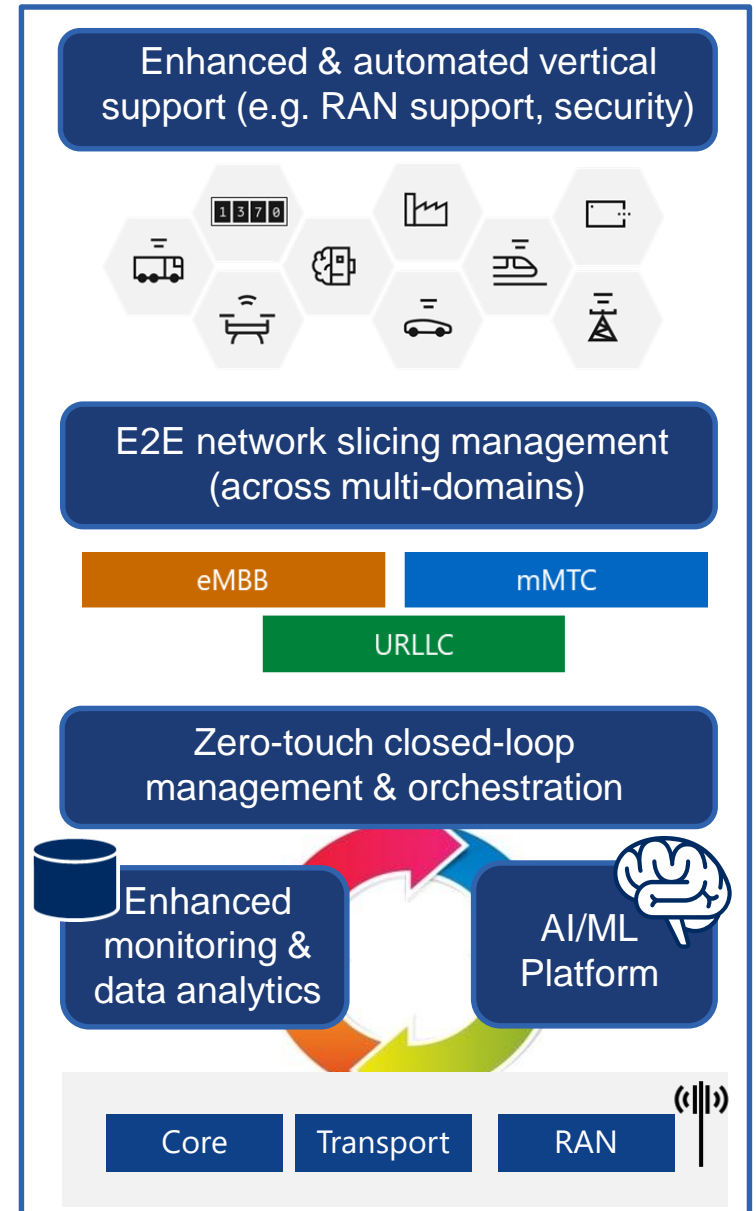
To empower vertical industries with an **AI-driven automated and shareable 5Growth Innovations**



Developed a **5G E2E service management platform (5Growth platform)** to provide zero-touch and automated provisioning and management of vertical services:

- **Vertical Slicer** as the vertical portal as interface with the verticals to get their service requests and map them to network slices;
- **Service Orchestrator** handles the service orchestration and manage their LCMs
- **Resource layer** manages the infrastructure resources and exposes resource abstraction to the higher layers.

5Growth Innovations



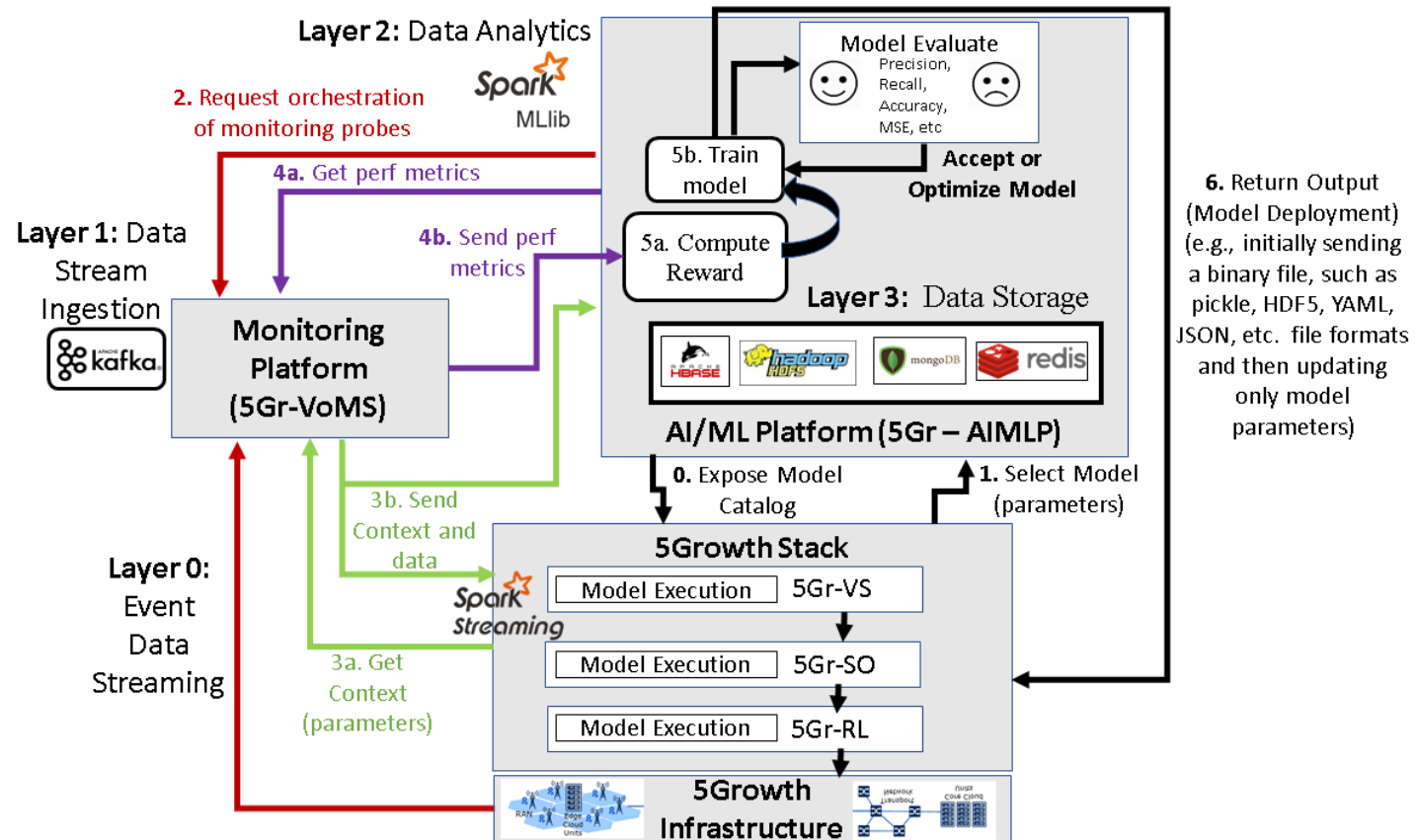
Closed-Loop Automation Framework

This framework aims to enable the **closed-loop automation for service Life-Cycle Management**.

Following the closed-loop design from the ETSI Zero-touch and Service management (ZSM) framework

These closed-loops include the process of

- Collecting monitoring data from the services and networks.
- Performing real-time data analytics for identifying events to handle.
- Making orchestration decisions for optimization/reconfiguration of the system.



Reference: Xi Li et al. "5Growth: An End-to-End Service Platform for Automated Deployment and Management of Vertical Services over 5G Networks", in IEEE Communications Magazine, vol. 59, no. 3, March 2021.

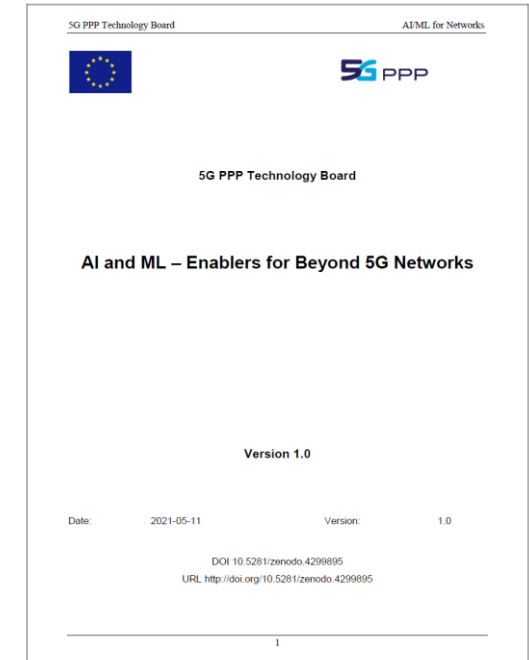
Use Cases of Applying AI/ML

Different use cases of applying AI/ML for 5G and B5G networks are described in the recently published 5G PPP AI/ML white paper

"AI and ML – Enablers for Beyond 5G Networks" (May 2021)

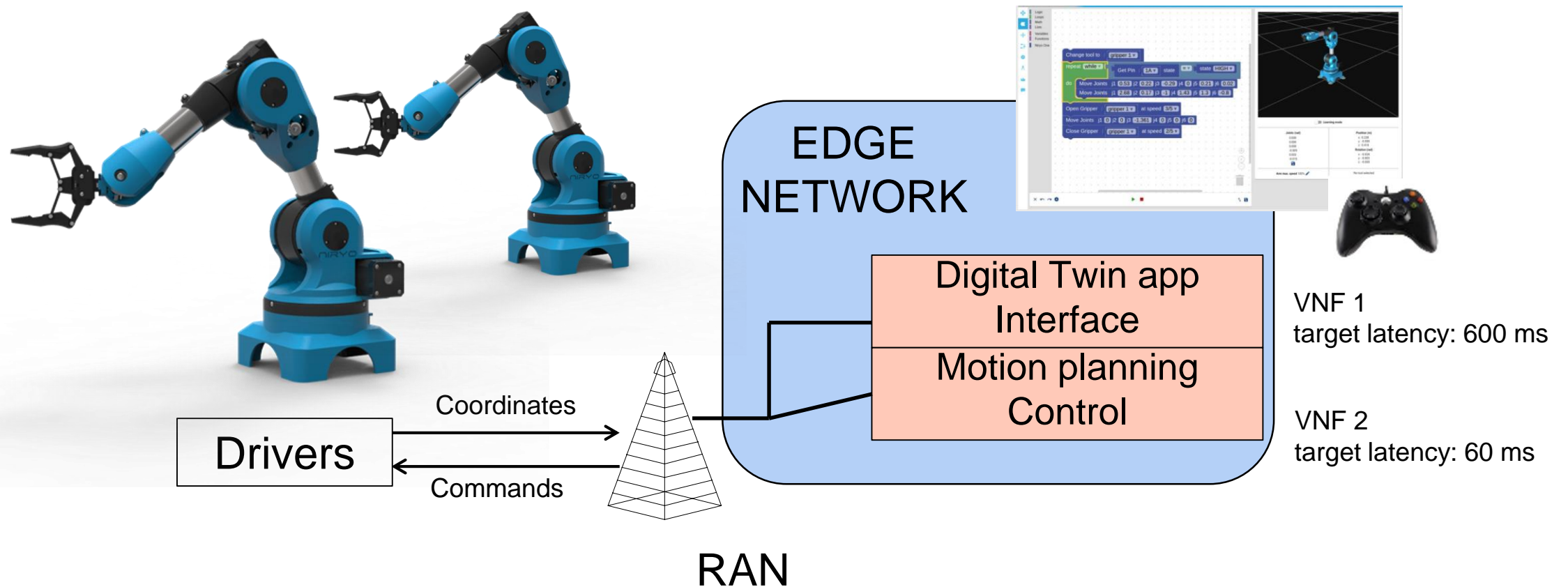
[online: <https://5g-ppp.eu/wp-content/uploads/2021/05/AI-MLforNetworks-v1-0.pdf>]

- Application of AI/ML in radio access: the slicing in multi-tenant networks, radio resource provisioning and scheduling.
- Application of AI/ML in transport networks: path computations, traffic management on the programmable switches, etc.
- Application of AI/ML for computing infrastructure: dynamic resource allocation for service function chaining in the NFV infrastructure, usually formulated as VNF-FG embedding problem
 - VNF placement: decide where the virtual infrastructure each VNF will be hosted and allocation of physical resources
 - Traffic steering: decide the network paths to transfer data between VNFs and bandwidth allocations
- Application of AI/ML for E2E slicing: automated E2E service assurance, resource allocation, automated scaling, slice isolation, etc.
 - Decide the required number of service VNF instances and corresponding virtual resource allocations in correspondence to dynamic changes of the service demands, in order to ensure the desired SLAs while compliant to given network constraints



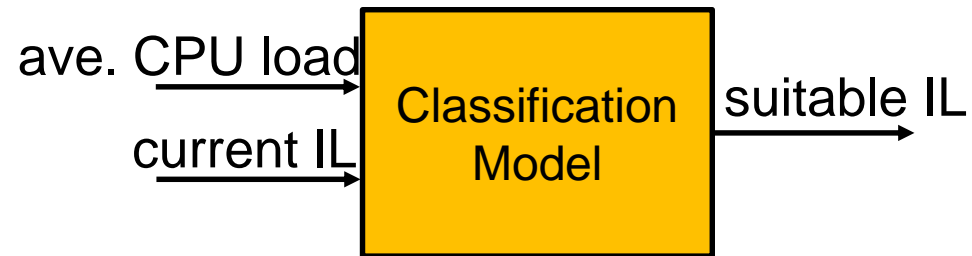
Industry 4.0 Digital Twin Use Case

- The evaluation considers a **Digital Twin** (DT) network service for a **Industry 4.0 scenario** considering multiple instantiation levels (ILs)



ML Model

- Random Forest classification model
 - Input: current Instantiation Level (IL) and average per-VNF CPU load
 - Output: IL to apply

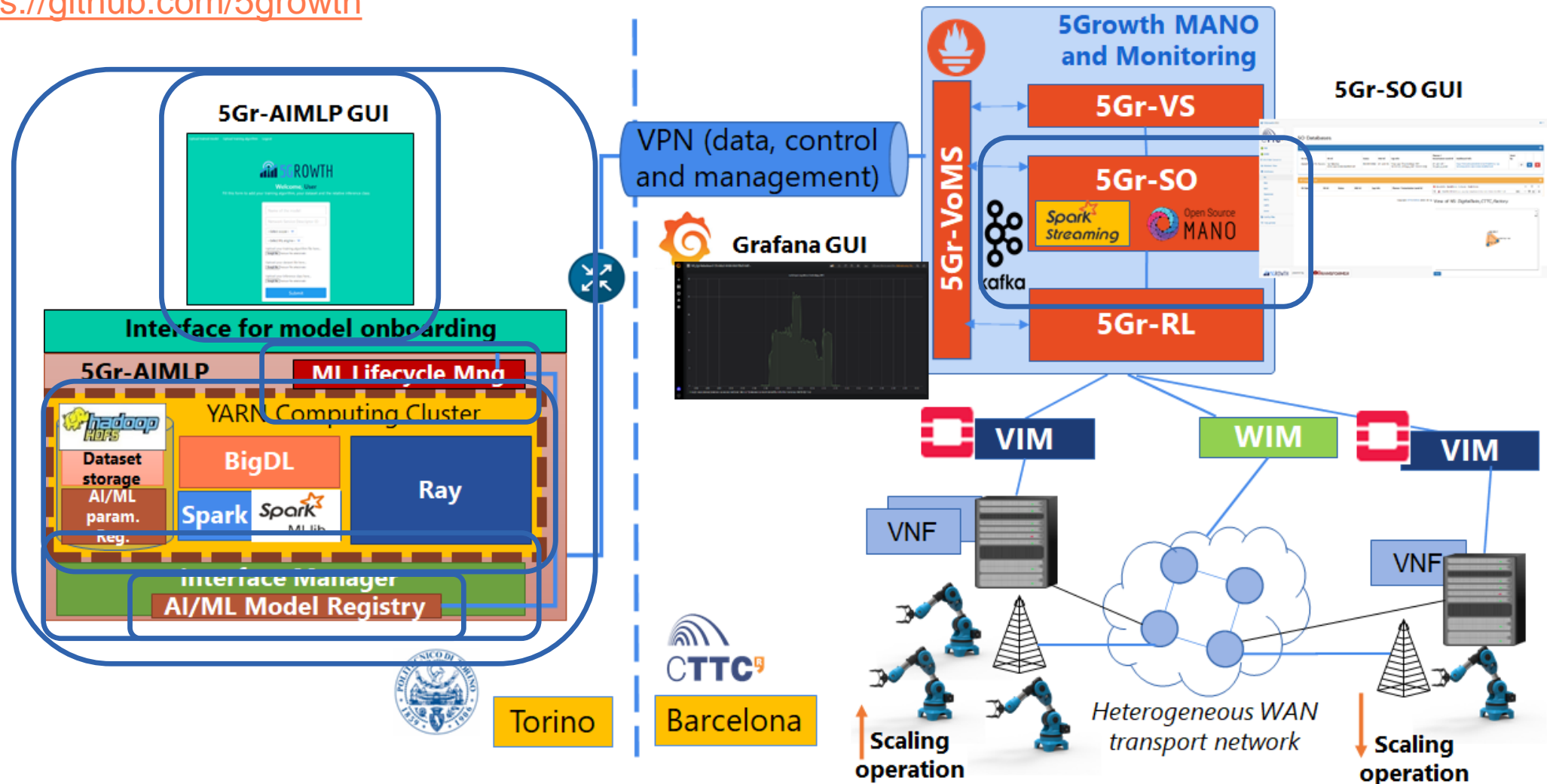


- Model training and inference model creation
- Future work is to explore the Reinforcement Learning tools to maximize a collective reward to ensure the latency requirement of different VNFs

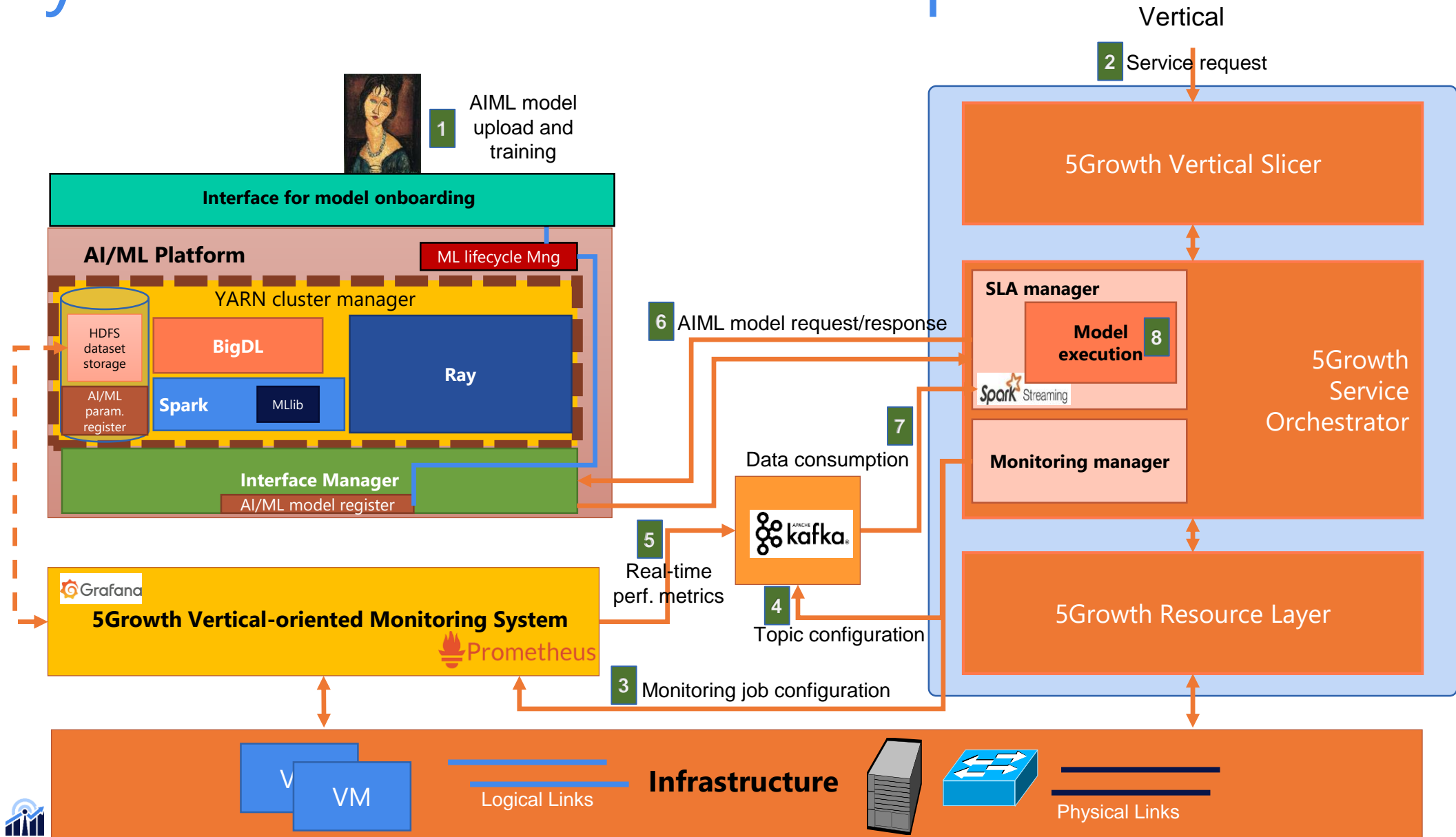
Deployed Proof-of-Concept

The software Implementation of the 5Grwth platform is provided as open source for R&D community

<https://github.com/5growth>



System architecture and operational flow



Testbed Result for the DT use case Service Instantiation

- **Used AI/ML model:** Random Forest Algorithm Trained offline
 - Training Dataset: 318K (20.2 MB)
 - Resulting model: 86.5KB (46.4 KB devoted to inference file)
 - However, configuring the pipeline is just a small part of the instantiation process, with VNF deployment being the most time-consuming step.
 - Overall, the impact of AIML-related operations on the whole instantiation process is limited to about 3.5% of the total average instantiation time measured at the 5Gr-SO (39.78 ± 3.24 s).

5Growth Building Block	Operation	Measured Time
5Gr-AIMLP	Dataset upload (offline)	10.3 ± 2.0 s
	Training (offline)	35.6 ± 2.5 s
5Gr-SO	Model/Aux file download (instantiation phase)	281.9 ± 20.4 ms
	AIML-related operation (instantiation phase)	1367 ± 40.9 ms



Testbed Result for the DT use case Service Scaling

Operation	Measured Time	AIML impact
Scaling out	26.72 ± 2.75 s	9.55%
Scaling in	24.54 ± 2.58 s	10.32%

Scale out represents the time to move from one to two instances of VNF2 and scale in, the opposite operation, from two to one instance



Key Take-Aways

- Novel design of AI/ML platform integrated into the 5Growth service management MANO stack, building a closed-loop service automation system enabling autonomous networks driven by AI/ML.
 - Proof-of-concept for a digital twin use case:
Less than 10% latency overhead
- The code implementation is released as open source for R&D community <https://github.com/5growth>
- Stay tuned for 5Growth updates on <https://5growth.eu/>
- Visit 5Growth demo videos on <https://www.youtube.com/5GROWTH>



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 856709.