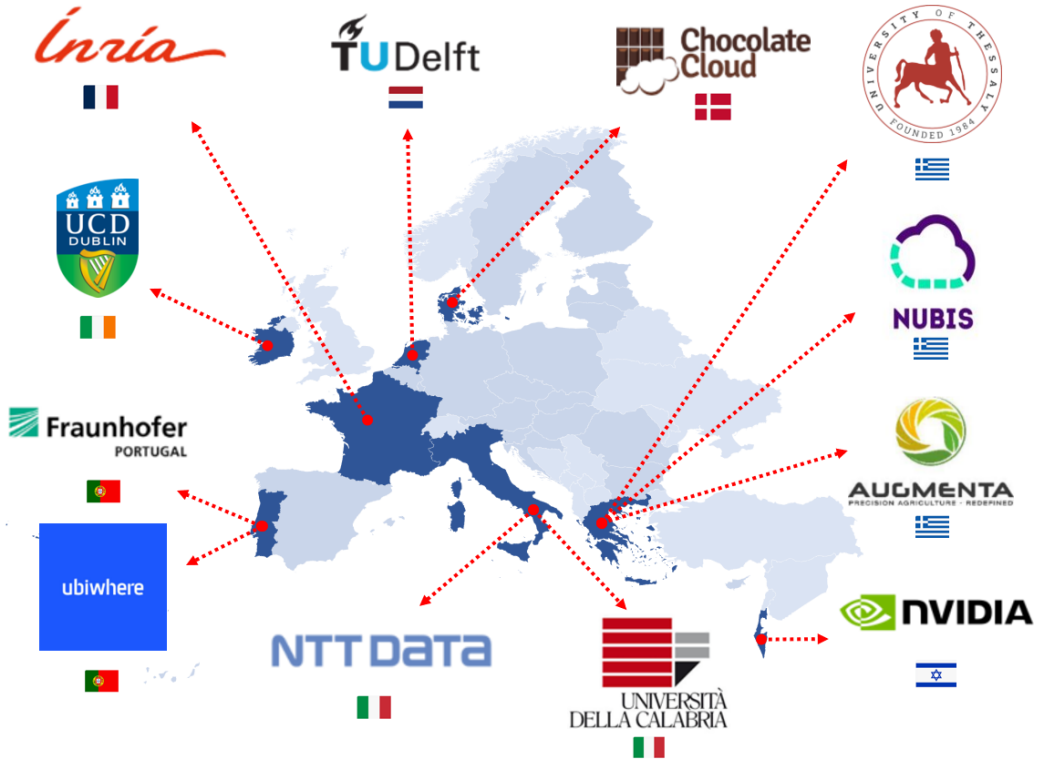
European Commission | HORIZON EUROPE

MLSysOps

MLSysOps

Christos Antonopoulos

University of Thessaly



Trends & challenges

Continuum systems

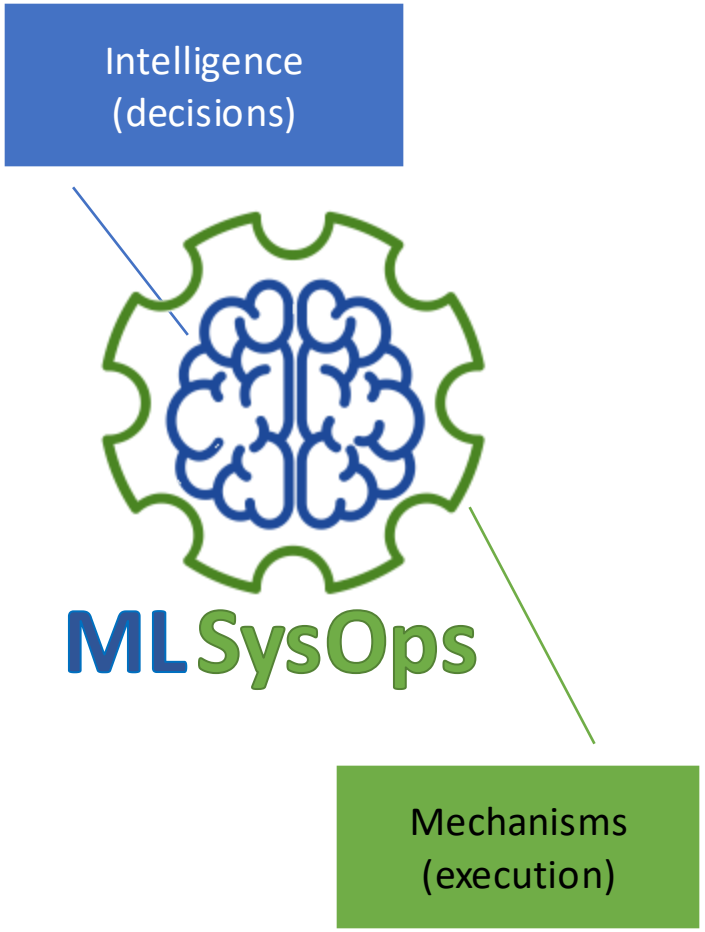
- Applications are moving outside the cloud and start involving nodes and resources toward or directly at the edge of the Internet
  - including powerful but also resource-constrained IoT devices
- Scale, heterogeneity, dynamics and complexity increases
- Practically impossible to monitor/manage by a human

*The vision of **autonomic computing systems**, which can manage themselves with little or no involvement of the application/system administrator, becomes more relevant than ever before!*

The MLSysOps project

*Autonomic system management and configuration in the **cloud-edge-IoT continuum** using **AI/ML methods***

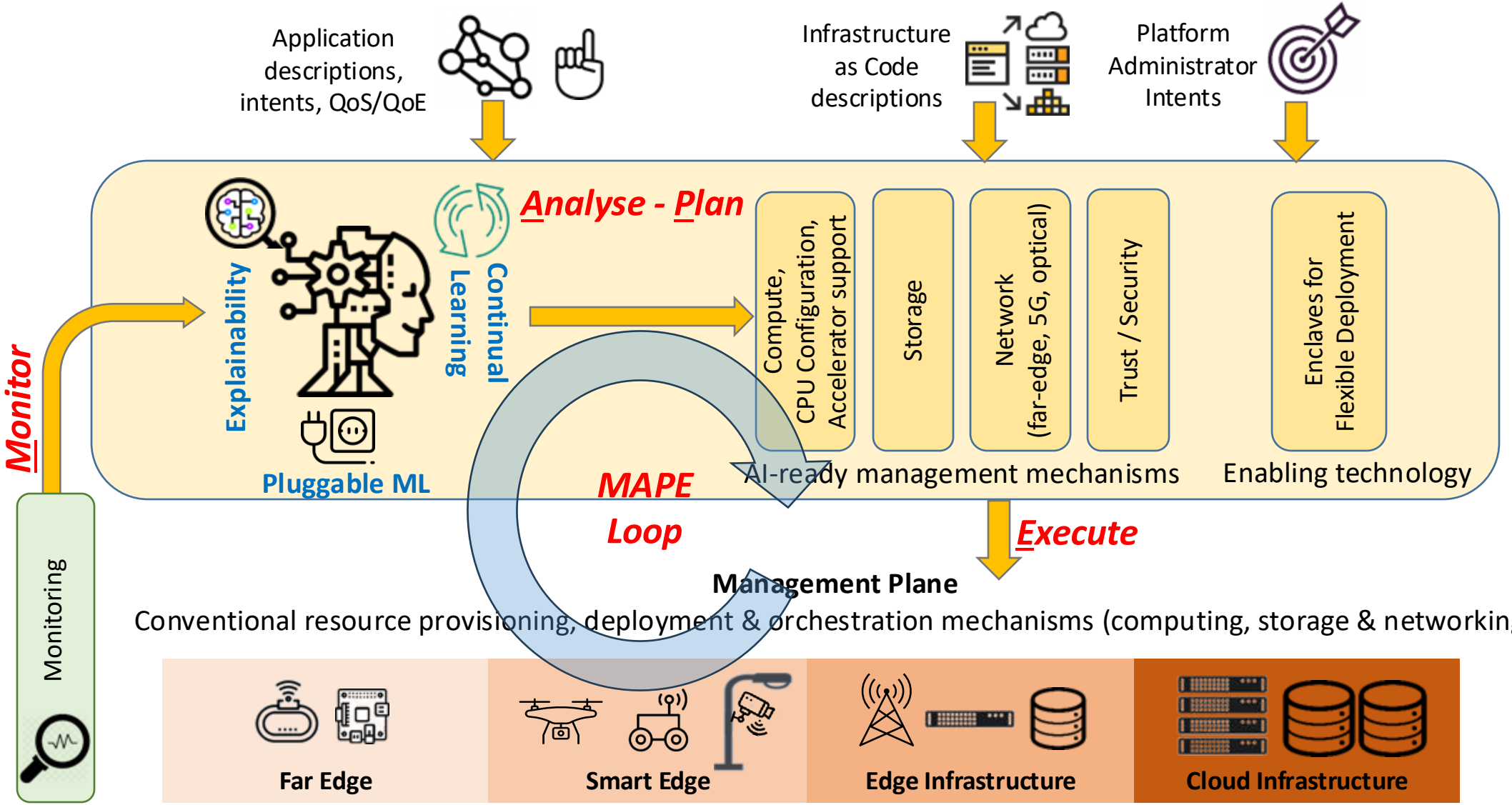
- Focus on modular, distributed applications
- Explore different management aspects
  - Deployment, computing, storage, communication/networking, trust
- Disassociate management from control
  - Develop AI/ML-ready (policy-neutral) mechanisms
  - Take decisions using suitable ML models
- Key AI/ML properties
  - Distributed / hierarchical approach
  - Continual learning / efficient model retraining
  - Explainability
- Extensive data generation and evaluation via real-world application testbeds, research testbeds and simulators



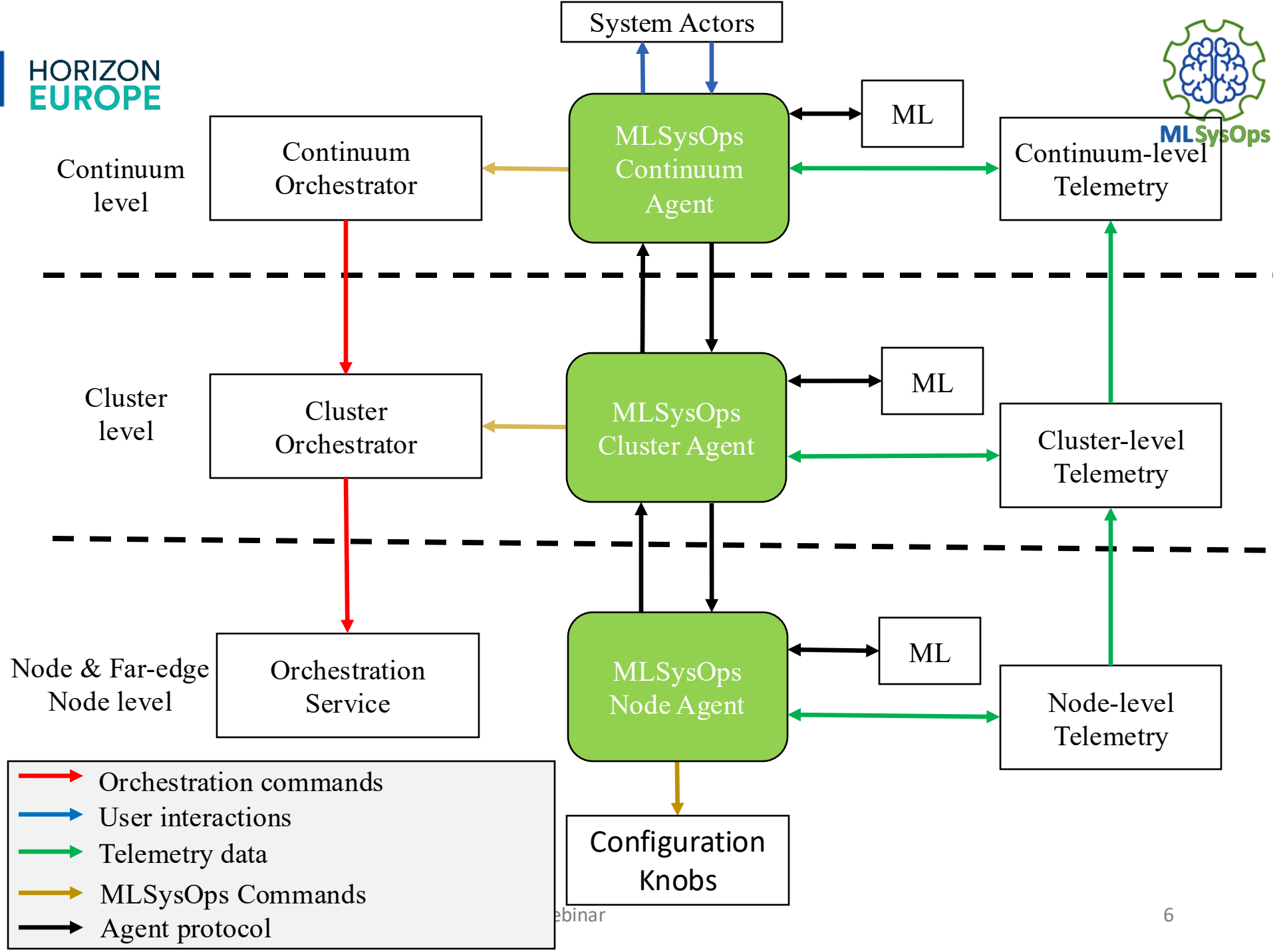
What?

- Open, AI-ready, agent-based framework for holistic, trustworthy, scalable, and adaptive system operation across the heterogeneous cloud-edge continuum.
- AI architecture supporting explainable, efficiently retrainable ML models for end-to-end autonomic system operation in the cloud-edge continuum.
- Efficient, flexible, and isolated execution across the heterogeneous continuum.
- Green, resource-efficient, and trustworthy system operation, while satisfying application QoS/QoE requirements.
- Realistic model training, validation and evaluation.

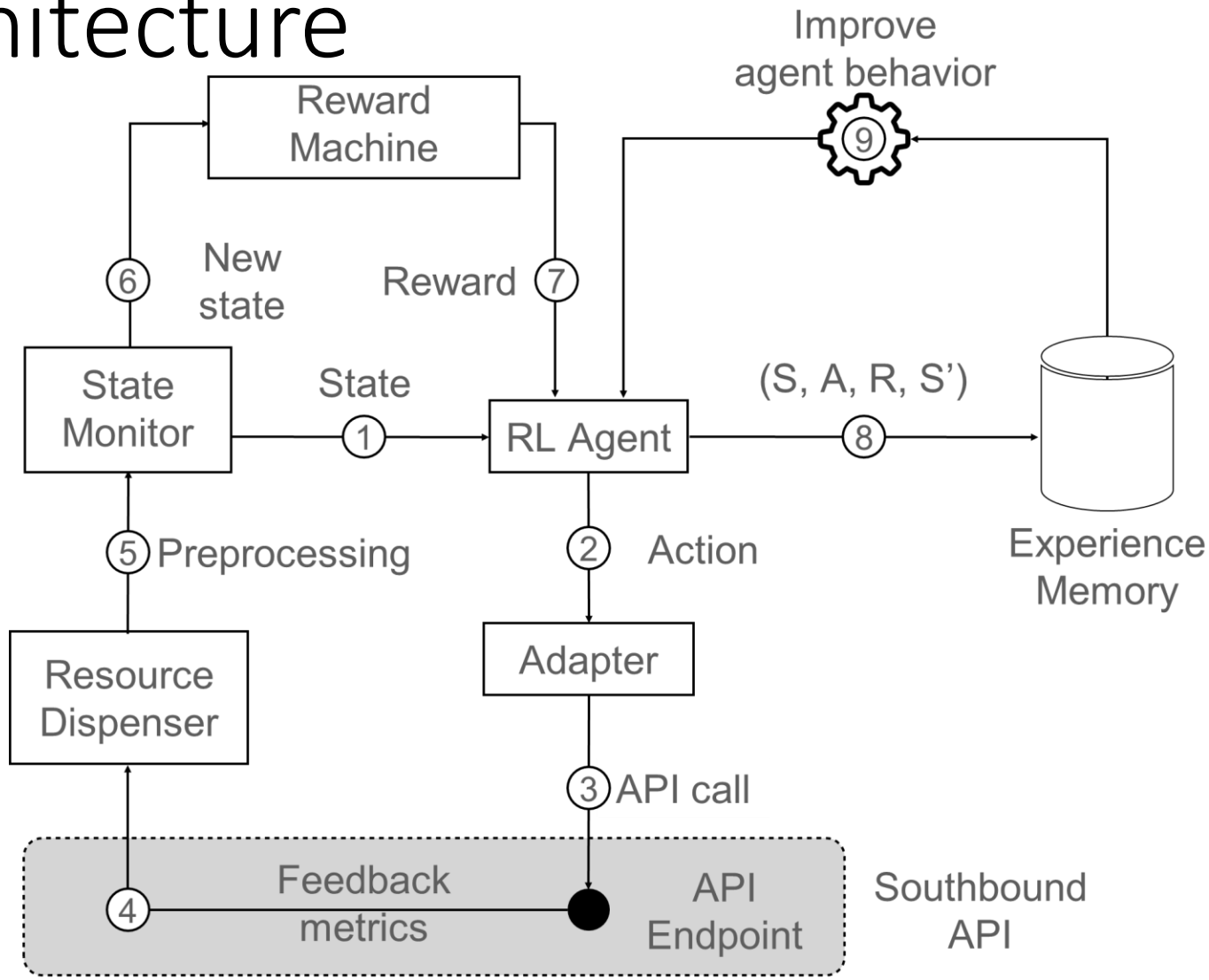
How? Concept


Application descriptions, intents, QoS/QoE
Infrastructure as Code descriptions
Platform Administrator Intents
Analyse - Plan
Explainability
Continual Learning
Pluggable ML
Monitor
Compute, CPU Configuration, Accelerator support
Storage
Network (far-edge, 5G, optical)
Trust / Security
Enclaves for Flexible Deployment
AI-ready management mechanisms
Enabling technology
MAPE Loop
Execute
Monitoring
Management Plane
Conventional resource provisioning, deployment & orchestration mechanisms (computing, storage & networking)
Far Edge
Smart Edge
Edge Infrastructure
Cloud Infrastructure

How? Hierarchy



ML Architecture



Two real-world application use cases

Smart Cities

Improve existing system via AI/ML-driven control to manage/configure application modules that can be deployed on smart lampposts and/or to datacenters.

Smart Agriculture

Improve existing system via AI/ML-driven control to jointly manage/configure the image processing application pipeline on a tractor and a companion drone.

Technical achievements

- Detailed requirements captured
- Framework architecture design completed

- First mechanism implementations
- First ML-based logic based on RL
- First agent implementations

- 5 simulation environments
- 2 real-world application testbeds
- 5 research testbeds

Coming soon…

- Workshop organization proposal (HiPEAC 2025)
  - In collaboration with other projects of the cluster

- Hackathon organization (late 2025)
  - Familiarize developers with the MLSysOps framework
  - Foster application & ML models development

- Open-source components of the MLSysOps framework

Coming soon…

- Workshop organization proposal (HiPEAC 2025)
  - In collaboration with other projects of the cluster

- Hackathon organization (late 2025)
  - Familiarize developers with the MLSysOps framework
  - Foster application & ML models development

- Open-source components of the MLSysOps framework

https://mlsysops.eu

Thank you!

@mlsysops

mlsysops

@mlsysopsproject

Newsletter https://mlsysops.eu/communication/