

Agentic Service Assurance for Cloud-Edge-Device Continuum

Turning AI Agents into Measurable SLO
Improvements (Real-World Operations)

Dr. Mohan Liyanage
mohan.liyanage@fh-dortmund.de

Seeking Consortium: WP lead/co-lead on benchmarking + agentic QoS loop.

Project EMULATE FH-Dortmund

<https://www.fh-dortmund.de/microsite/smartergelab/projekte/emulate.php?loc=en-US>



The Gap: AI Predicts But It Doesn't Act

Problem & Motivation



- **The Reality:** Cloud-Edge-Device Continuum is dynamic, heterogeneous, and often a "black box".
- **The Problem:** Current AI stops at prediction. It alerts but doesn't fix.
- **The Consequence:** Recurring SLO violations (latency/throughput) that require manual intervention.
- **Our Solution:** Closing the loop.

Agents that Observe → Reason → Act autonomously.

The "Safety-First" Agentic Loop

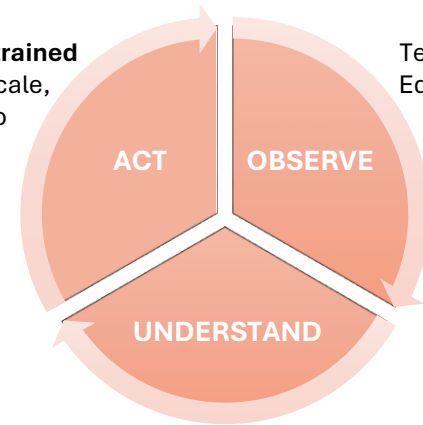
Project Idea



We add QoS & Energy value to **ALL** AI-Agent proposals.

- **The Problem:** Distributed AI Agents fail when the network jitters or energy is constrained.
- **Our Solution:** We provide the "Service Assurance Layer" that sits *underneath* your AI.
 - **Proactive:** Short-horizon forecasting to prevent upcoming SLO violations.
 - **Energy-aware:** Optimize SLO while considering energy/renewable constraints.
- **The Benefit:** We enable **your** agents to run with guaranteed **QoS** and optimized **Energy**, regardless of grid/network fluctuations.

Policy-Constrained Mitigation (Scale, Shift, Steer) to restore SLOs via tool API.



Telemetry from Edge, Network, & Your App.

Detect anomalies & localize bottlenecks under uncertainty.

Outputs & Validation



Benchmarking & Evaluation Framework

The Asset: We bring a working **Distributed Scenario Orchestrator** (TRL 5/6) to the consortium.

Scenario Library: Includes reproducible edge scenarios (load bursts, faults, mobility) already validated on critical **Automotive (V2X) and Drone** applications.

Standardization: Common definitions for SLOs and KPIs to ensure comparable results across partners.



Agent Tool / Actuation APIs

Unified Interface: Standardized APIs for safe execution of **Scaling, Shifting, Steering, and Configuring**.

Policy-Constrained: Built-in safeguards to ensure agents operate within defined operational boundaries — hardened through **3 years of industry collaboration**.



2-Stage Validation (Digital Twin → Pilot)

Stage 1 (Lab): Controlled experiments and Digital Twins driven by our existing dataset of real-world edge traces.

Stage 2 (Real World): Pilot-grade demonstration supported by the **Smart Edge Lab team (20+ experts)**.

Focus: Realistic constraints, stakeholder KPIs, and measurable benefits in a Cloud-Edge-Device Continuum / "Apply AI" setting.

Who We Need to Complete the Loop

Partners Sought



We are looking to join/form a consortium with:

- **Coordinator / Experienced Horizon lead:** To manage proposal, leadership, WP governance, exploitation plan.
- **AI-agent R&D leader(s):** Planning, tool-use, multi-agent coordination, robustness/explainability.
- **Telco / Infrastructure Provider:** To provide the real-world playground (requirements, constraints, pilot/testbed access).
- **Application Partner (Verticals):** Use-case owners (e.g., Automotive, Health, Industry 4.0) to define success criteria.
- **AI Sector Partner(s):** Use-case ownership, success criteria, end-user validation.
- **Optional:** Digital-twin / experimentation partners for scalable, repeatable evaluation.

Our Expertise

Smart Edge Lab: Core Competencies



Digital Twins & Emulation

End-to-End Emulation:

Digital-twin-enabled environments for controlled event generation and reproducible evaluation scenarios.

Controlled Scenarios:

Capability to simulate noisy measurements, time-varying workloads, and black-box component behavior.

AI-Driven Diagnostics

Root-Cause Localization:

QoS bottleneck detection under uncertainty using analytics-driven localization.

DNN-Aware Cost Models:

ML models for per-layer compute, energy, and data-transfer costs to support optimal partitioning and offloading decisions.

Energy-Aware Control Loops

Service Assurance:

Real-time monitoring combined with closed-loop corrective actions for the cloud-edge continuum.

Energy-Aware Mitigation:

Decision-making policies for workload shifting, elastic scaling, and renewable-aware ranking.

Consortium Role: We are ready to **Lead/Co-Lead Work Packages** on **Benchmarking & Evaluation** and **Agentic QoS Integration**.

Thank You

Project EMULATE FH-Dortmund

