17:15–18:00 | AI for NaaS and NaaS for AI – The Next Frontier of Intelligent Networking

WENYU SHEN

Senior Manager & Principal Architect

docomo Business

YUTA YAMAGISHI

System Architect

docomo Business

JOAN ESPIN GARCIA

Director of Network Platforms and Solutions Architecture

colt

ABHISHEK SINGHAL

Industry Principal

Infosys

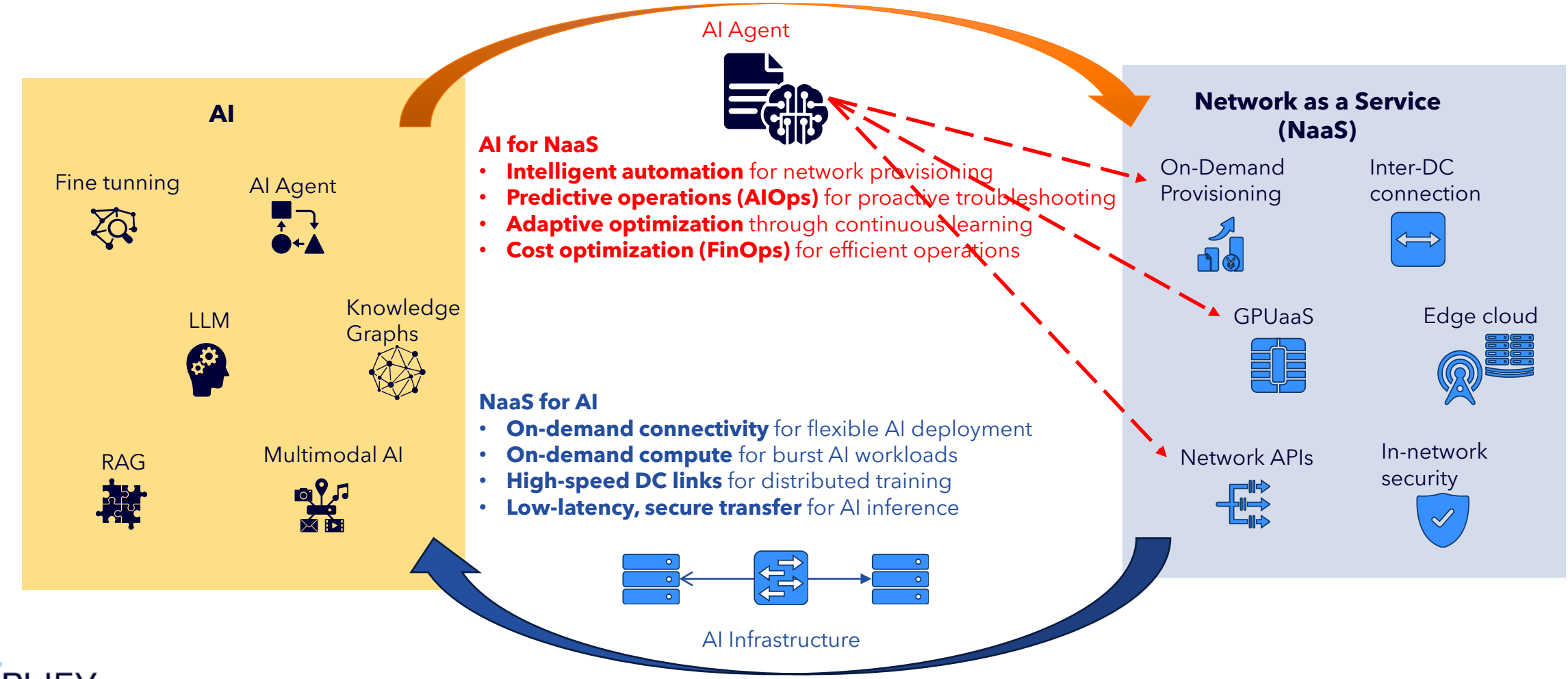
BARTOSZ MICHALIK

Principal Architect

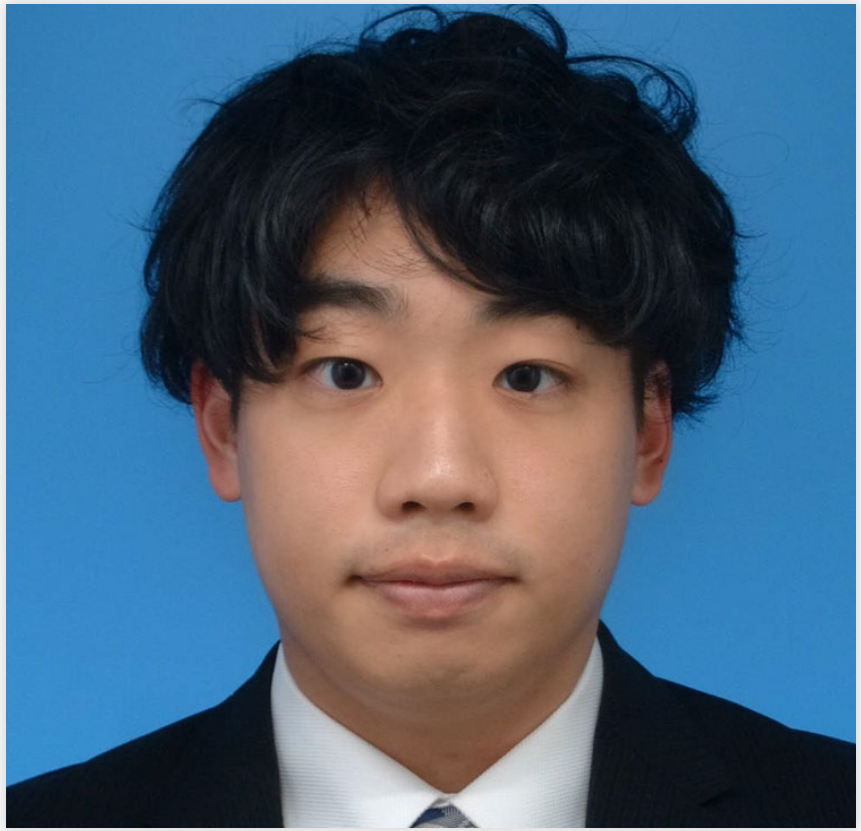
amartus

When AI Meets NaaS

- As **AI and NaaS converge**, their synergy is redefining digital infrastructure design, operation, and monetization.
- This session explores both dimensions of this transformation — **how AI empowers NaaS**, and **how NaaS enables AI**.



AI for NaaS



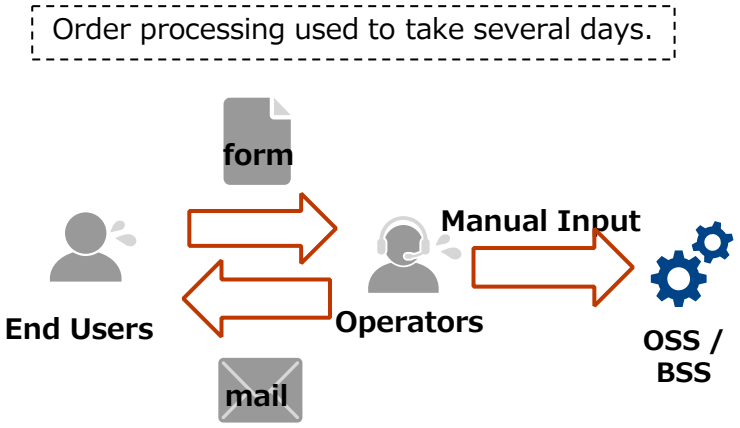
Yuta Yamagishi
System Architect
NTT DOCOMO BUSINESS

From Implementation to Standardization: Realizing AI for NaaS

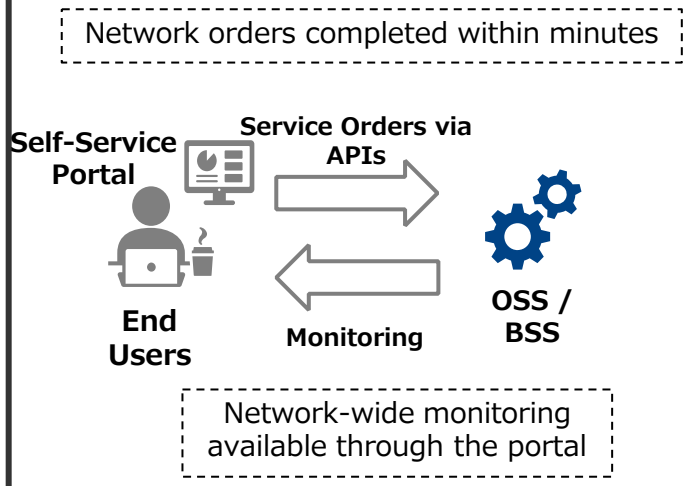
The Emergence of AI Agents in NaaS Services

- With the evolution of NaaS, network operations are shifting from "**leave it to others**" to "**manage it yourself**."
- The advent of NaaS has made **user-driven network operations** a reality. Now, with the introduction of **AI-Enhanced NaaS**, even **non-expert users** can manage networks **intuitively and effortlessly**.

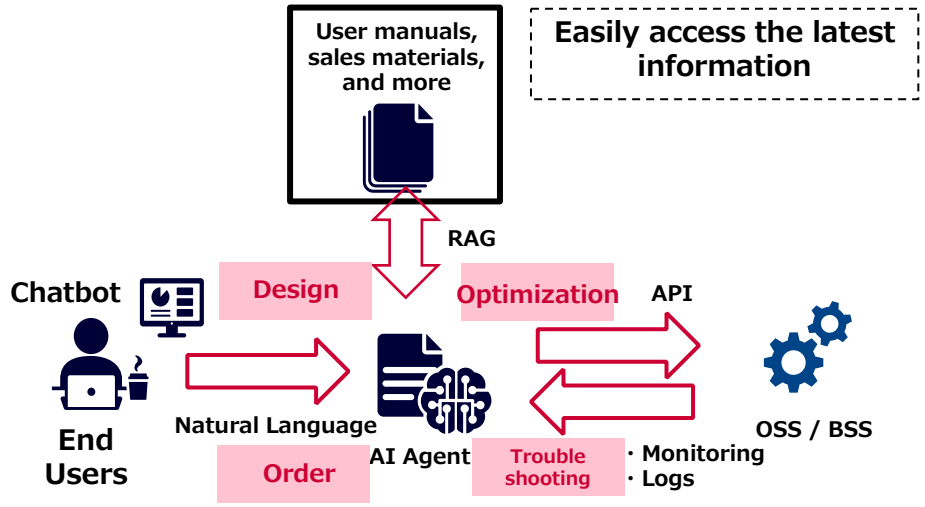
Legacy Network Services



The Current State of NaaS



AI-Enhanced NaaS



By enabling operation through natural language, even non-experts can use NaaS effectively.

AI Agent Use Cases in NaaS

Use Cases

Design

Input: Expected traffic, locations, applications, scale, etc.

End Users

Output: Generated topology, service options, and configuration steps.

Optimization

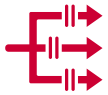
Propose configuration or contract changes to improve performance and enhance security.

End Users

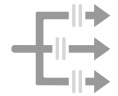
AI-Enhanced NaaS

Operation

Place service orders— for example, order new VPNs or modify security configurations via interactive chat.



NaaS API



Other APIs

Troubleshooting

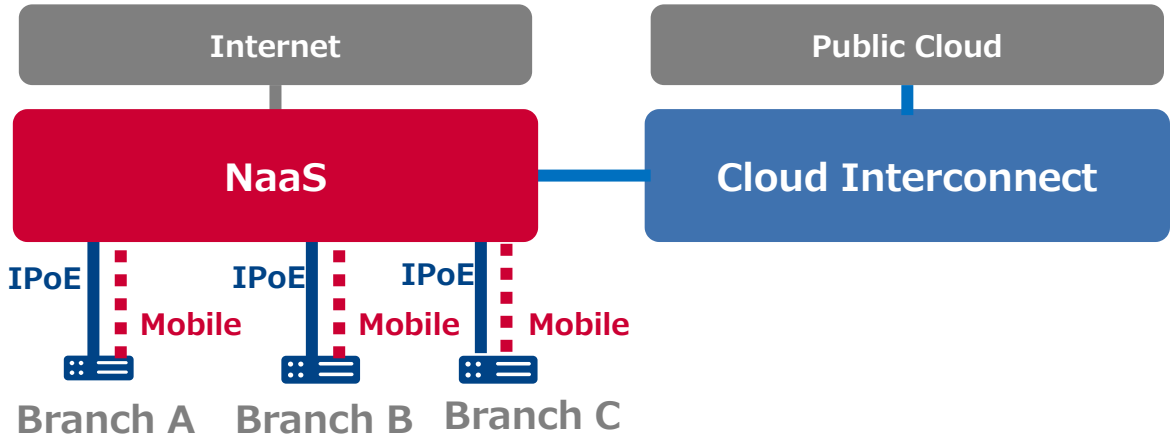
Intelligent root cause analysis based on a non-expert user's description

End Users

[Troubleshoot] Detecting and Resolving Network-wide Issues

Demo Scenario

- Gain an integrated view of NaaS operational status.
- Guide users through NaaS configuration updates.



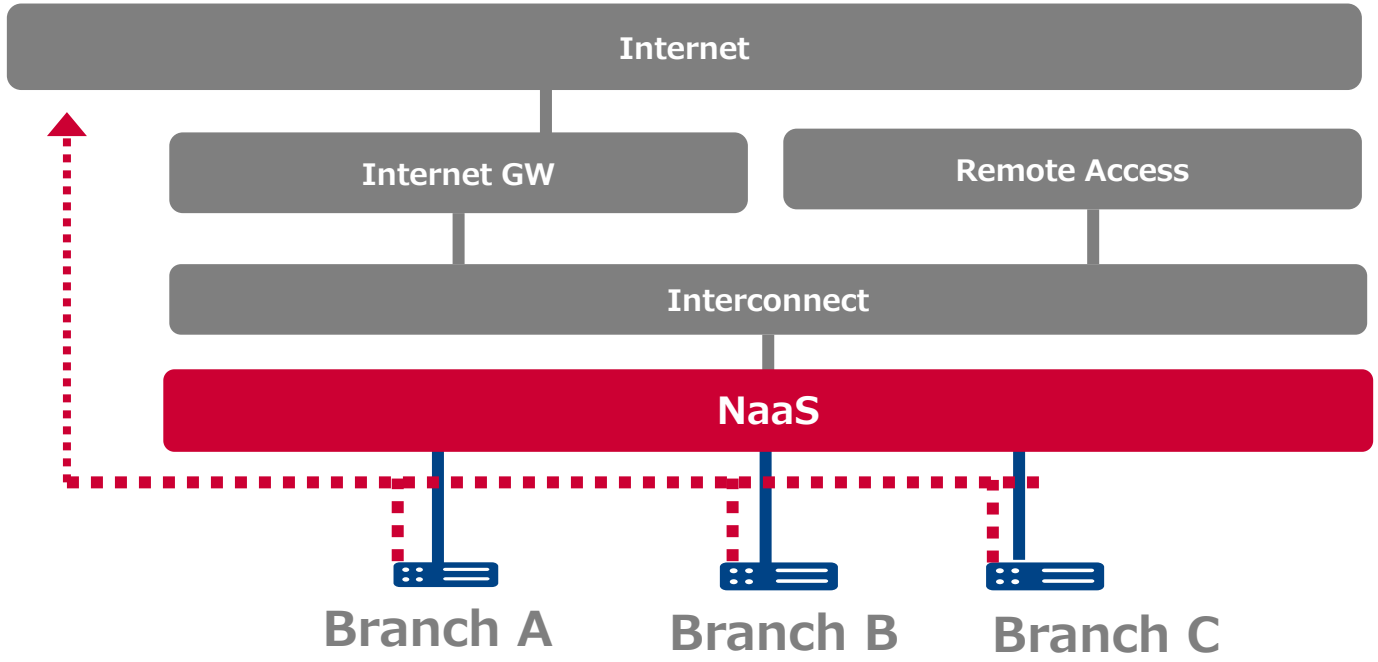
The AI agent integrates diverse NaaS environments, including inter-site VPNs and cloud connections. It autonomously troubleshoots issues, enabling non-network experts to easily operate and manage their corporate networks.

[Optimization] AI-Driven Network Optimization & Auto Configuration Updates

Demo Scenario

- Detect traffic congestion caused by specific applications.
- Assist in scheduling local breakout routes to optimize network performance.

Enable breakout routing for designated destination traffic.

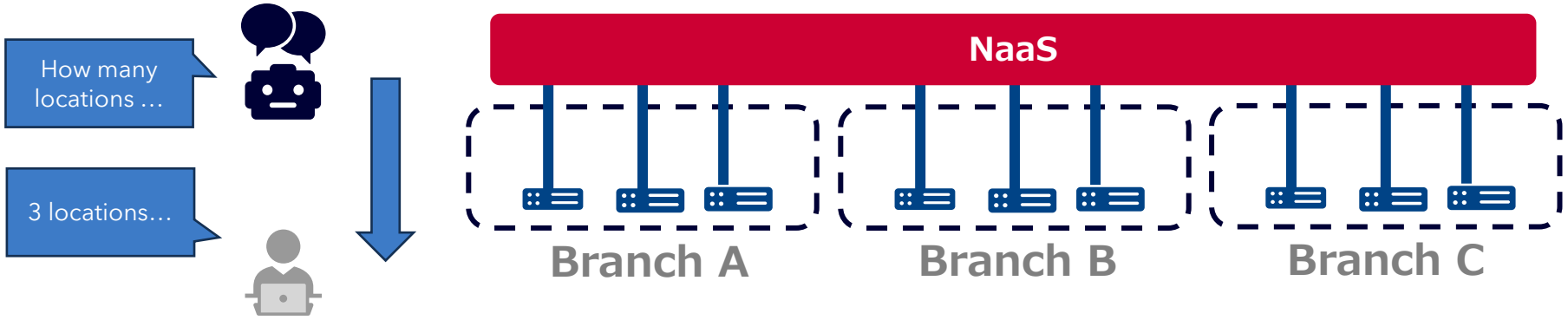


Analyzes traffic and applies optimal settings for better NaaS performance.

[Design] AI-based Network Design Support and Visualization

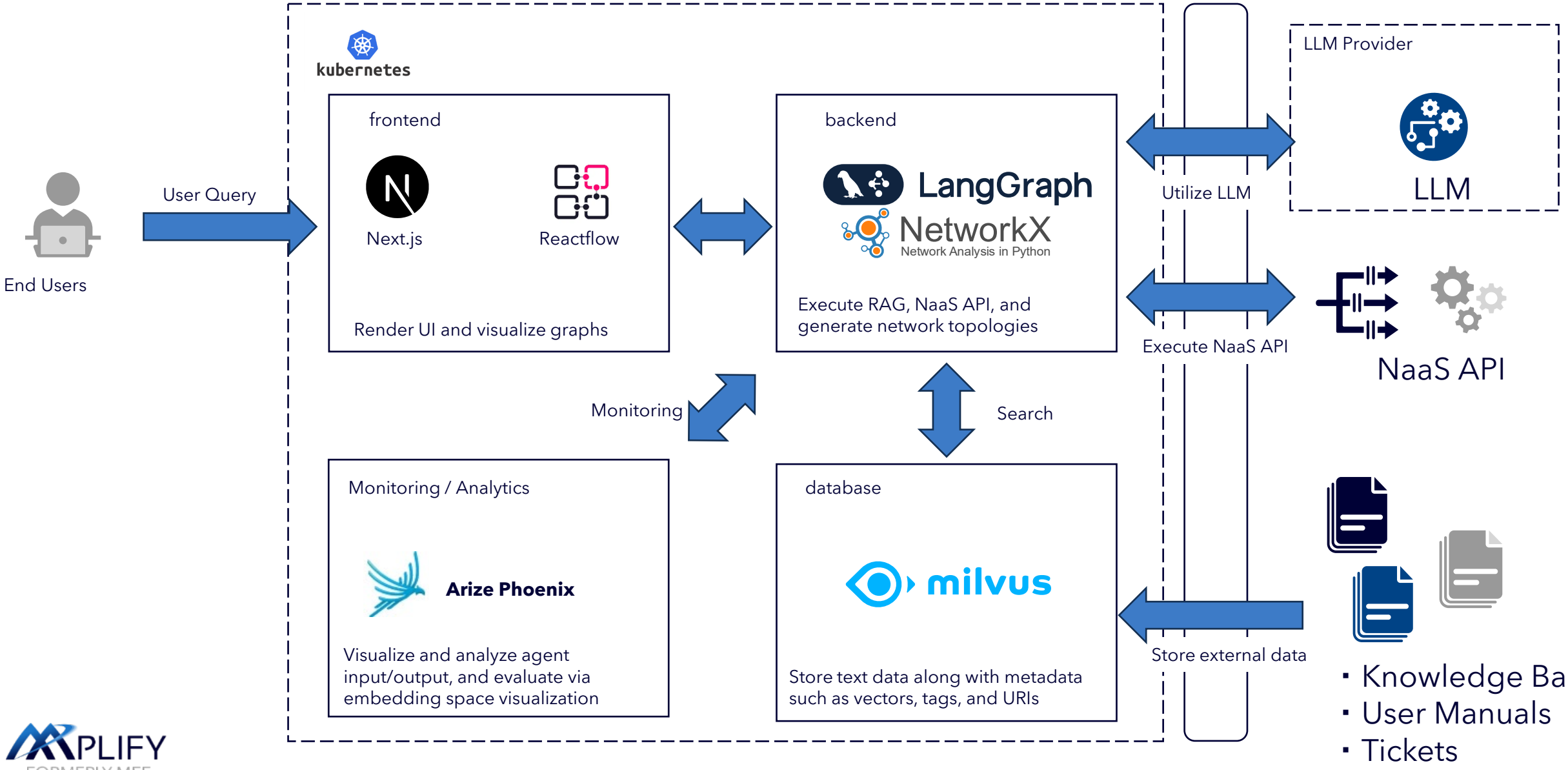
Demo Scenario

- An AI agent gathers user requirements while assisting with network design on the NaaS platform.
- The agent visualizes the appropriate network topology based on the gathered requirements.



The AI agent guides users through an interactive flow to design networks using NaaS resources.
It visualizes the optimal configuration and explains the design, encouraging NaaS adoption and network expansion.

System Architecture

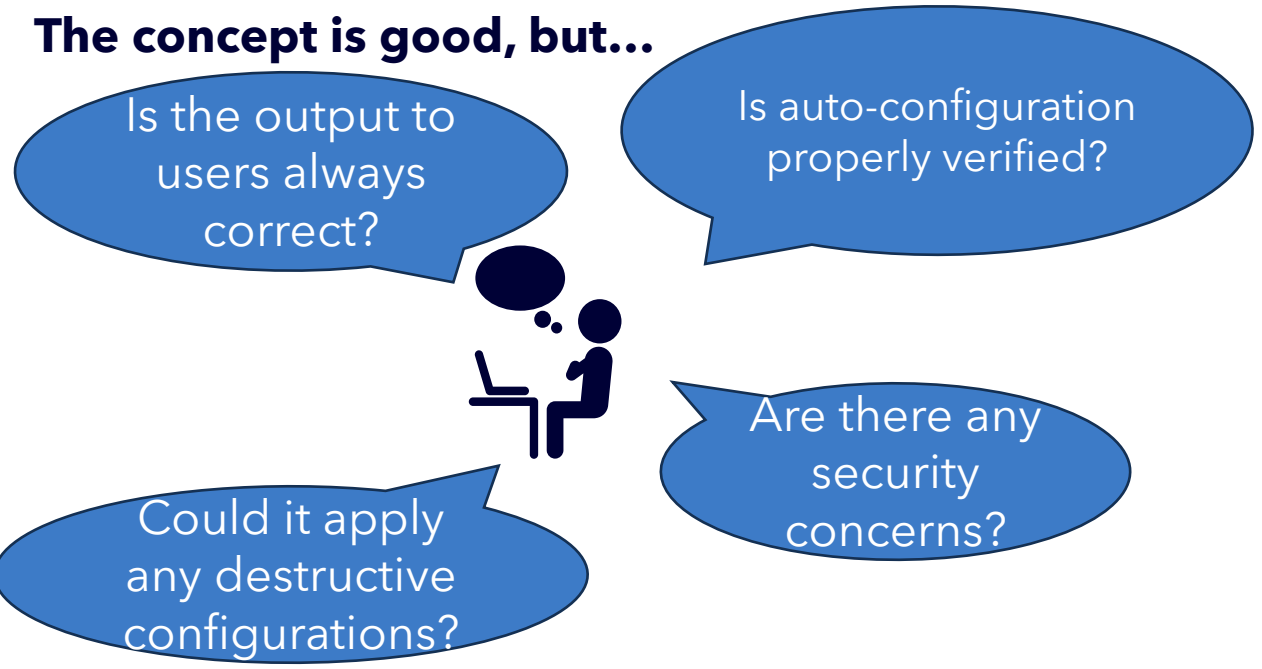


Focusing First on Internal User Development

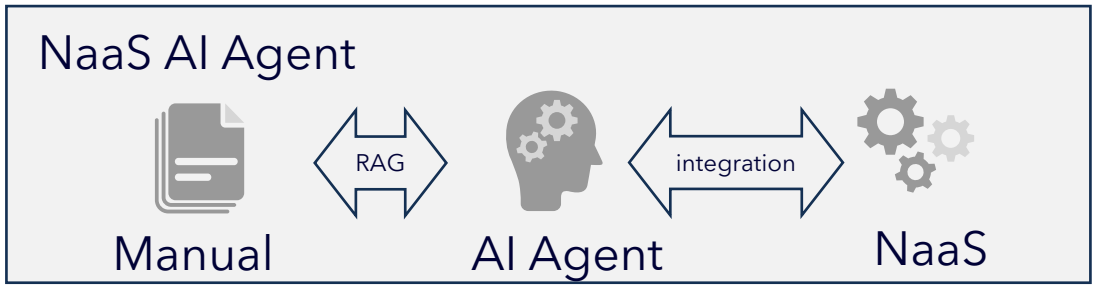
- Initial end-user application development exposed several commercial-level challenges.
- To address them, we proudly kicked off an internal PoC to support sales and operator teams, driving higher accuracy and a more robust system design.

External Launch for End Users

The concept is good, but…



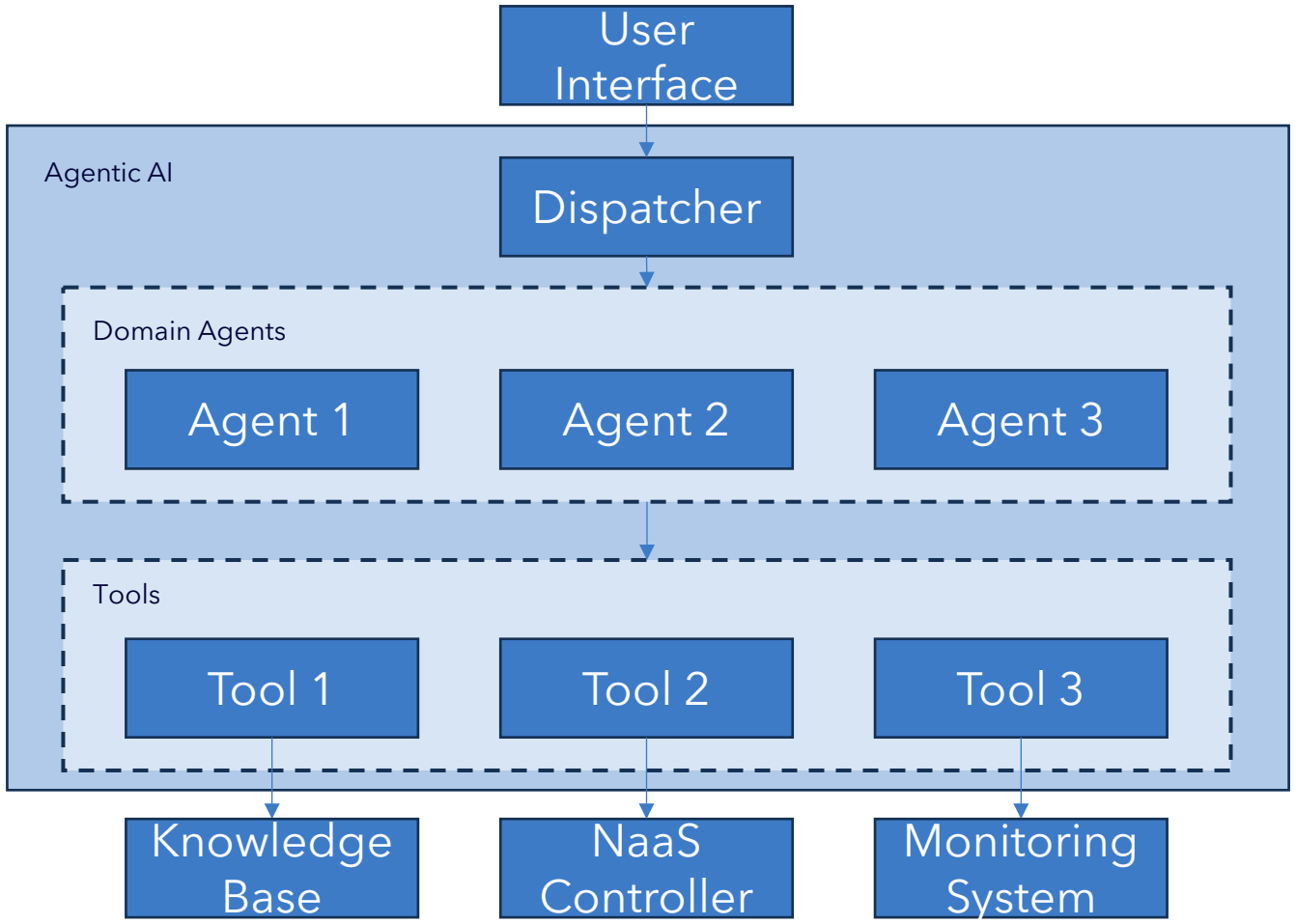
Internal Launch for Sales Team



Improving proposal efficiency and reducing learning costs through AI Agent support

Standardization of Agent Architecture

Example architecture of NTT DOCOMO BUSINESS



Discuss with Silent Comet, generalize, and contribute to the white paper.

Blueprint
by Silent Comet

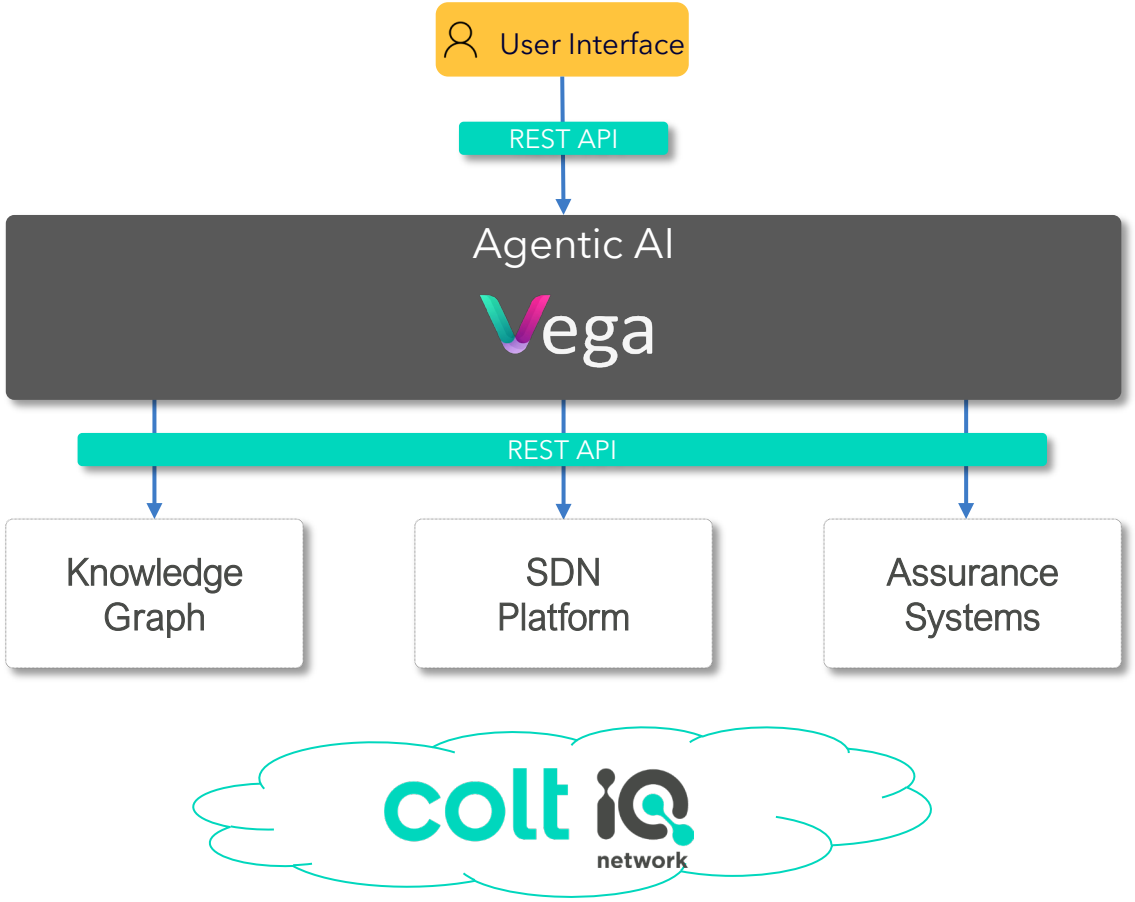
Joan Espín García
Director Network Platforms and Solutions
COLT

360° User Experience

Redefining UX via Declarative Network Intelligence

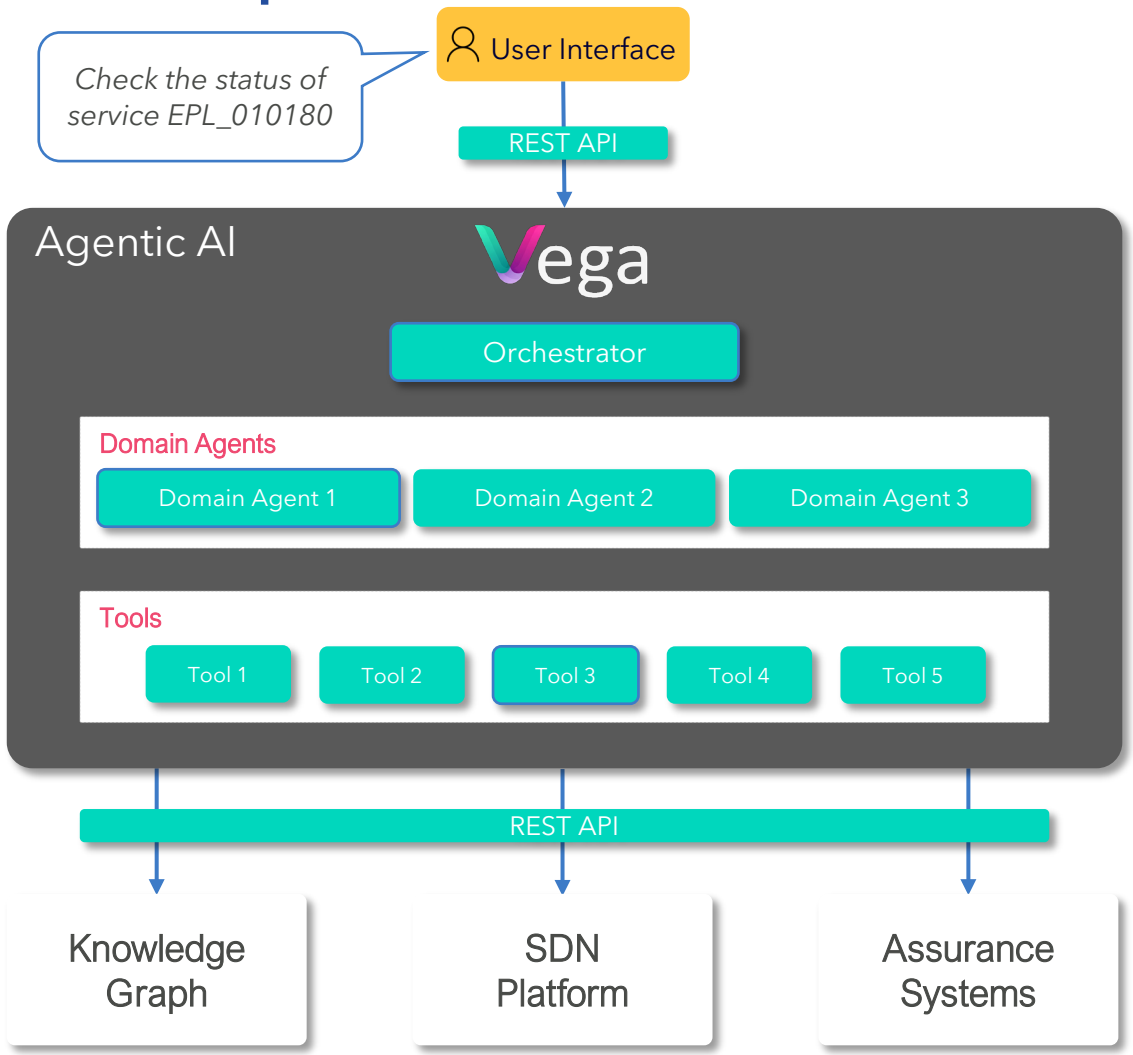
Agentic AI vision at Colt

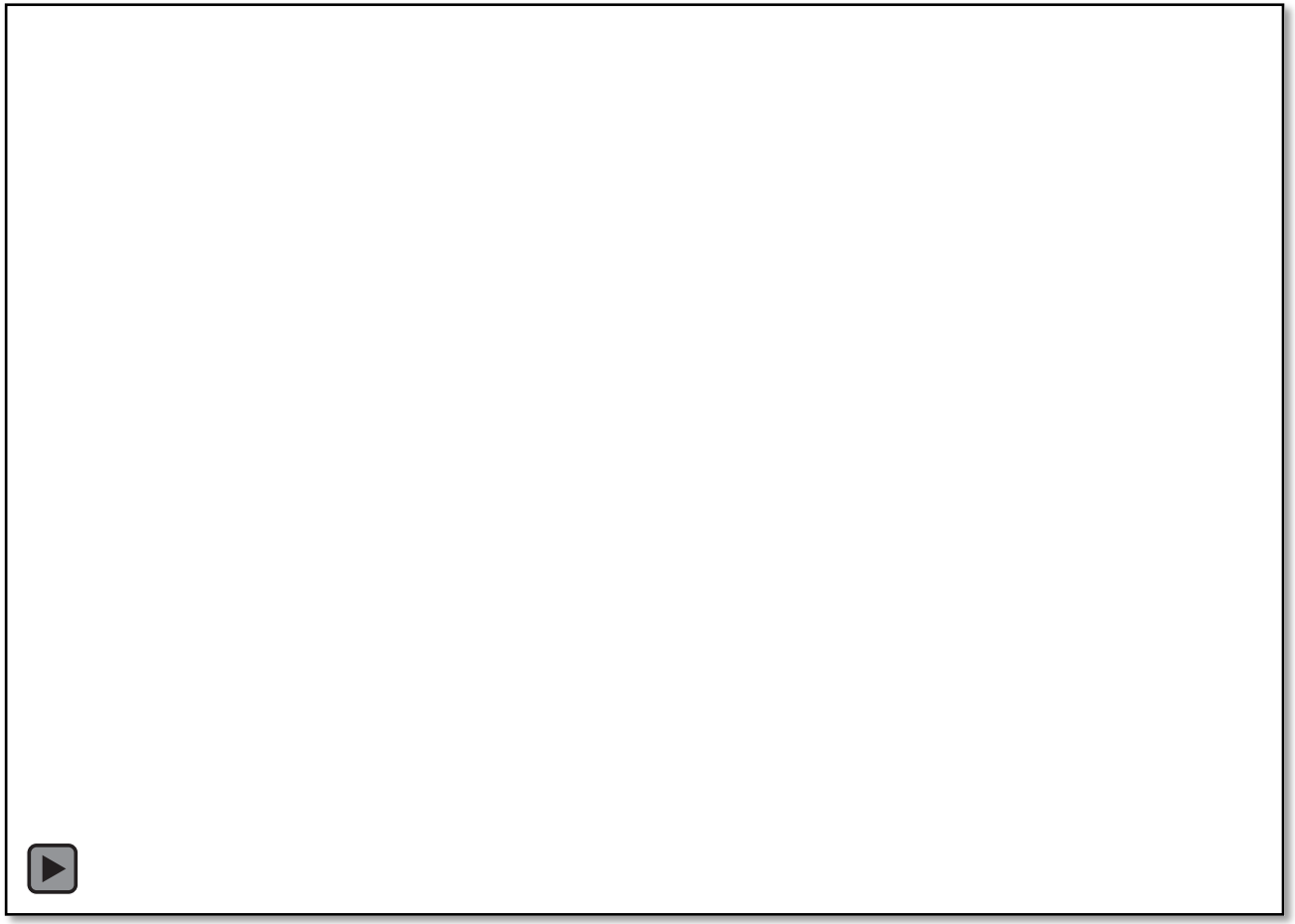
*"At Colt, we strive to make network interactions **Simple**, **Secure** and **Intuitive**"*



Vega introduces a new abstraction layer that translates business needs into network actions through intelligent automation

Example use case - Troubleshooting

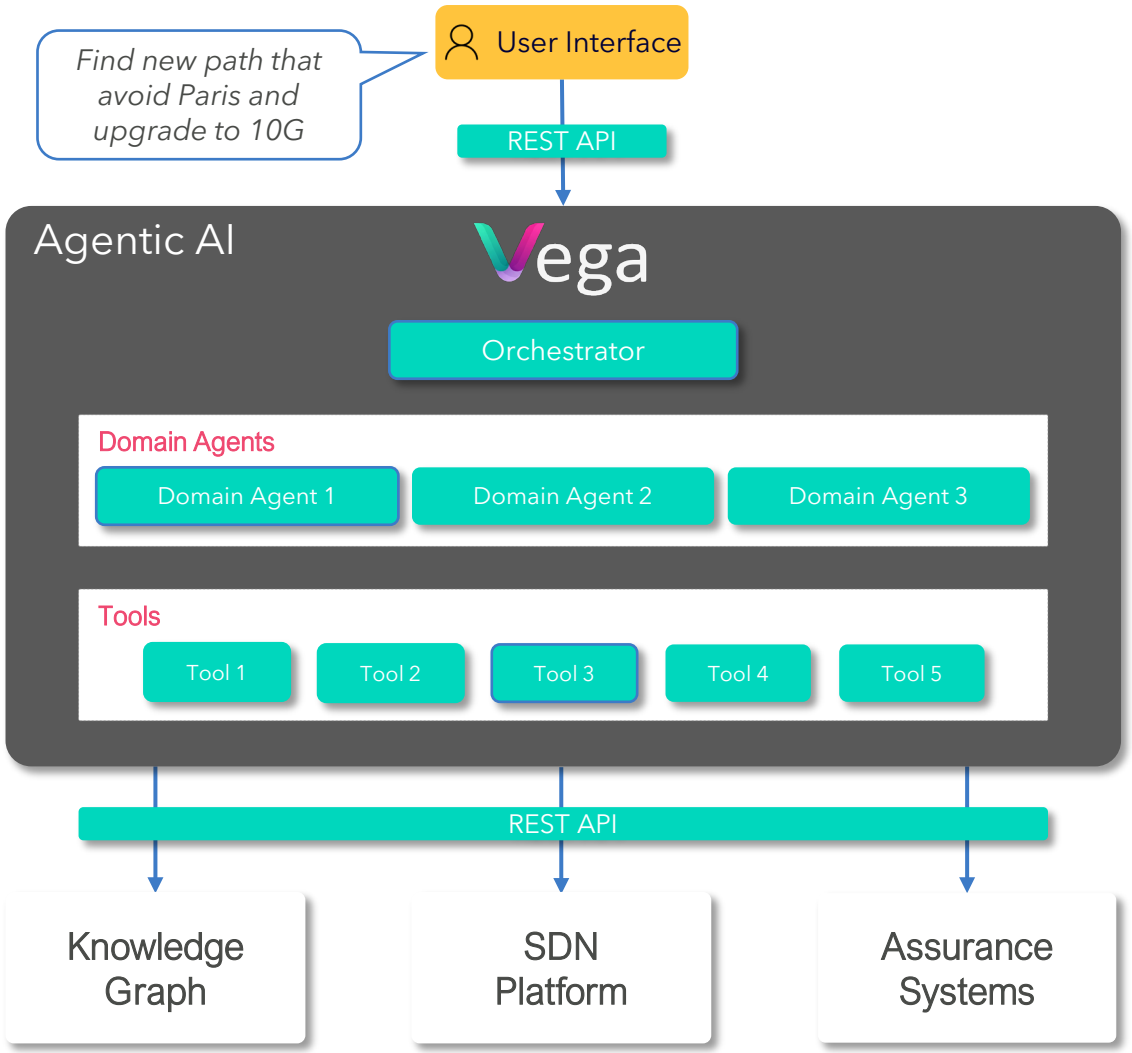


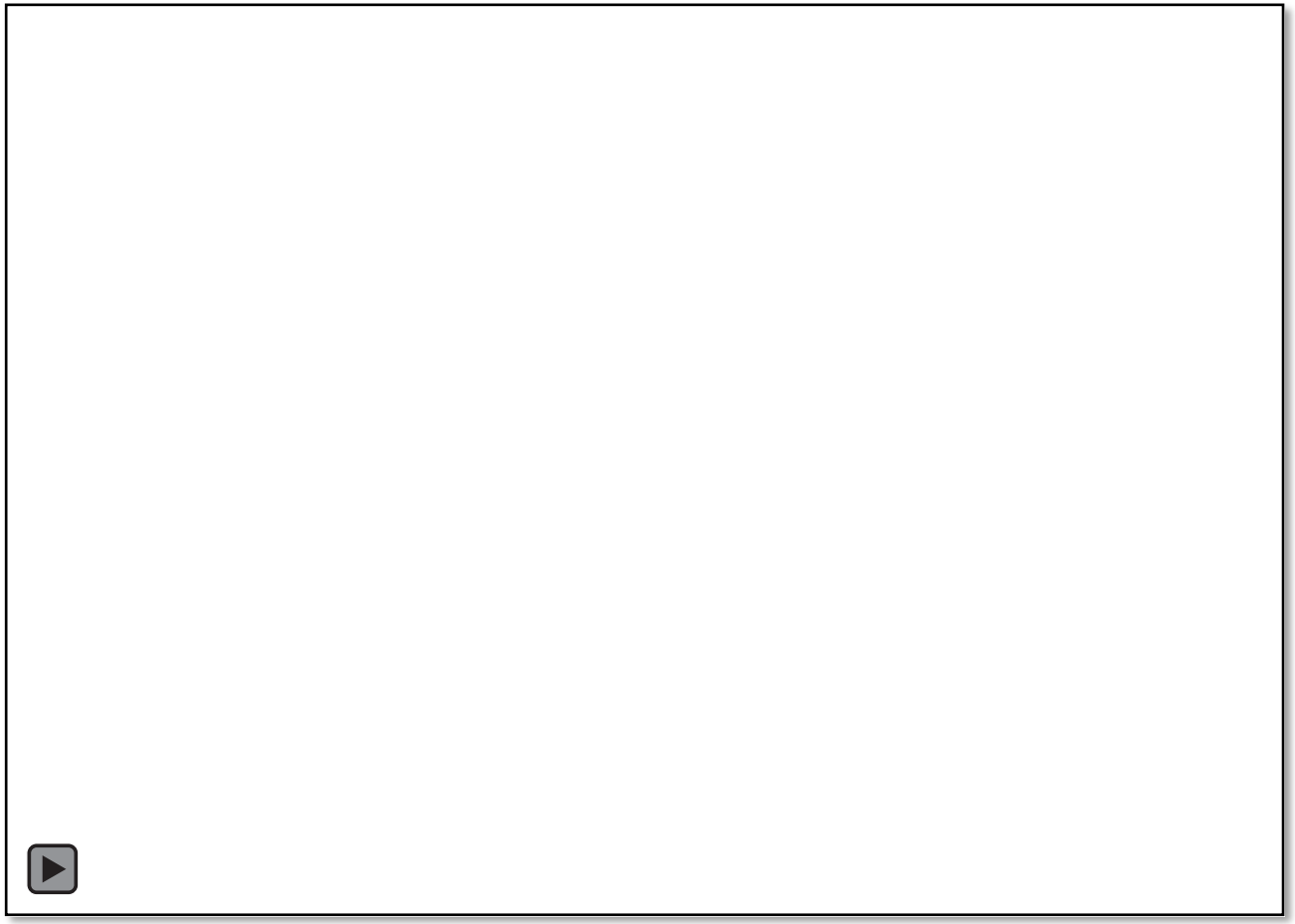


colt iQ network

AMPLIFY FORMERLY MEF

Example use case - Service recovery





Bartosz Michalik
Principal Architect
Amartus

Multi-agent LLM System Blueprint for Telecoms

AI adoption traps

Undefined ROI No success criteria or baseline KPIs. Six months in, stakeholders ask "Did this work?" with no data to answer.

Data Overreach Months building knowledge lakes instead of starting with a set of curated documents that solve one problem.

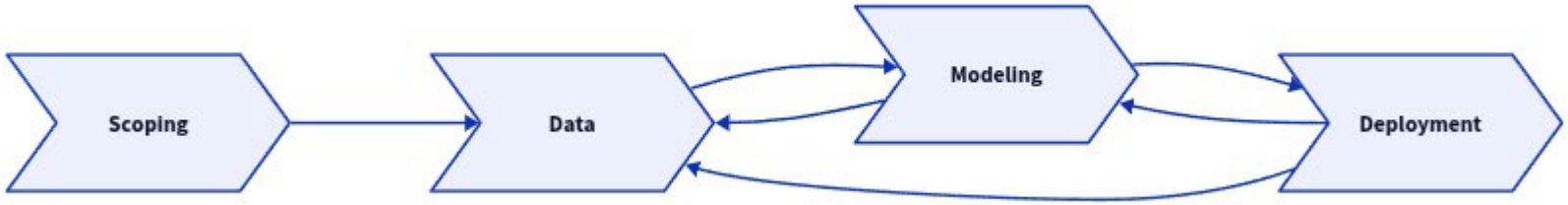
Fragile Prototypes Notebooks lack version control and quality gates. First production incident reveals no rollback or audit trail.

Uncontrolled Costs No token budgets or cost attribution. Cost per interaction exceeds manual processing, eroding the business case.

Disconnected ML Existing ML assets (anomaly detection, forecasting) remain siloed. Teams rebuild capabilities that already exist.

Two-lenses framework

Execution
lifecycle



Architectural
viewpoints

Lifecycle phases populate viewpoints → Insights drive next iteration

Combined effect: faster delivery, auditable decisions, controlled costs

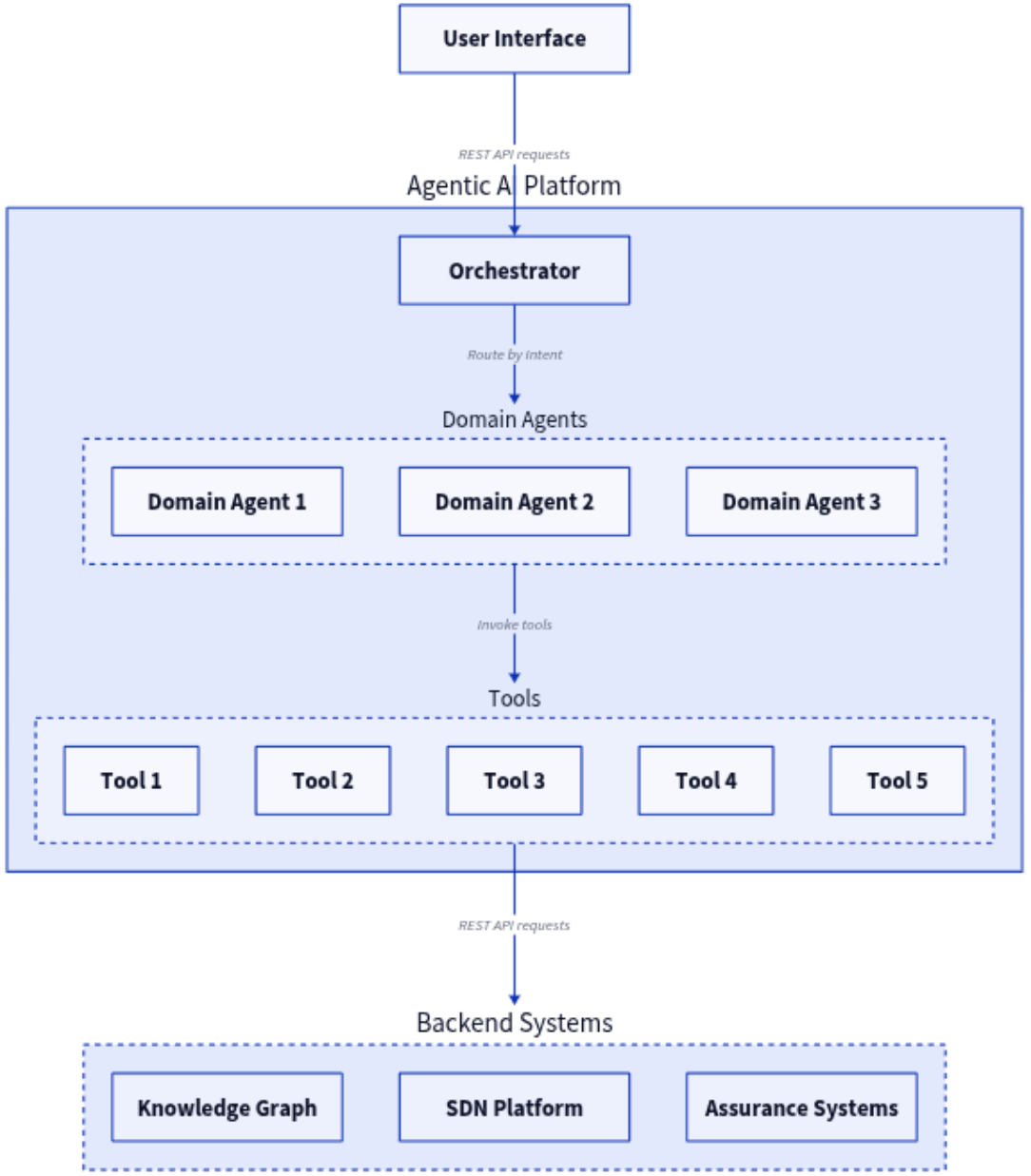
Reference architecture

Orchestrator enforces governance and routing across all use cases

Domain Agents execute the four-phase lifecycle within specialized domains

Tools centralize integrations, enabling component reuse across implementations

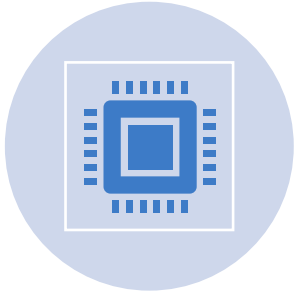
Validated by **NTT Docomo Business** and **Colt Technology Services**



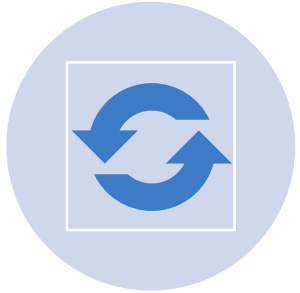
Take aways

Start narrow: One process, 3-5 KPIs, curated datasets

Governance enables speed: Controls = faster approval



Design for reuse: Three-layer architecture amortizes investment



Evolve incrementally: Start with 3 viewpoints, enrich from telemetry

NaaS for AI

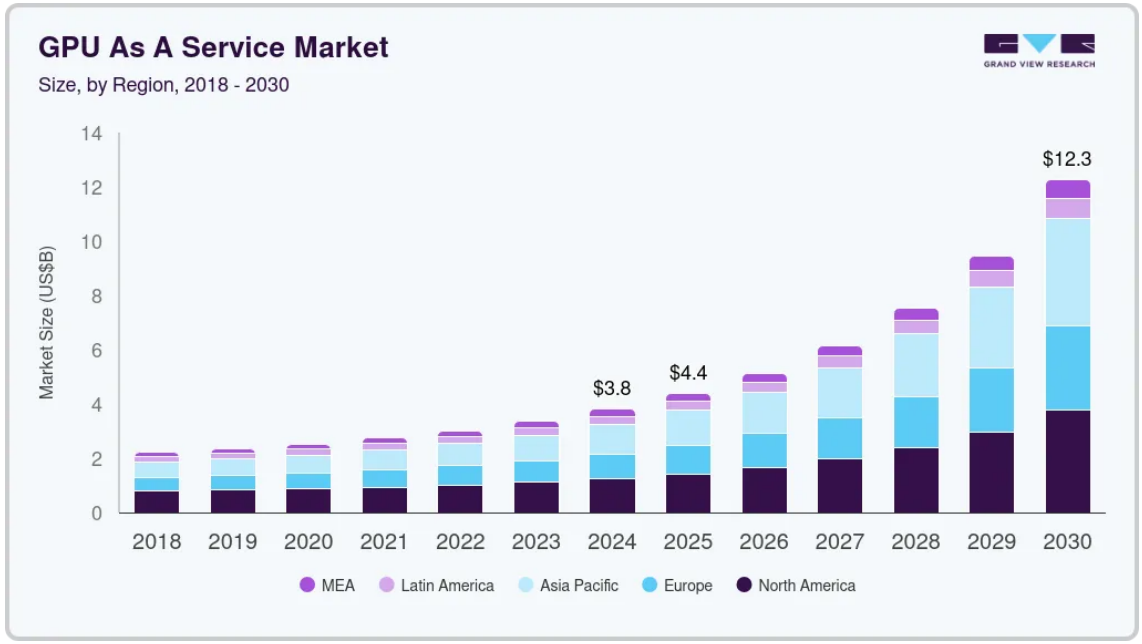
Wenyu Shen
NTT DOCOMO BUSINESS
Senior Manager & Principal Architect

GPUaaS: Powering the AI Era with On-Demand Compute

GPUaaS Market Overview

- The global GPUaaS market size is projected to reach USD 12.26 billion by 2030, growing at a CAGR of 22.9%.
- Neocloud --- a next-generation cloud model designed specifically for AI workloads, shifting from general-purpose compute to GPU-centric, service-oriented infrastructure.

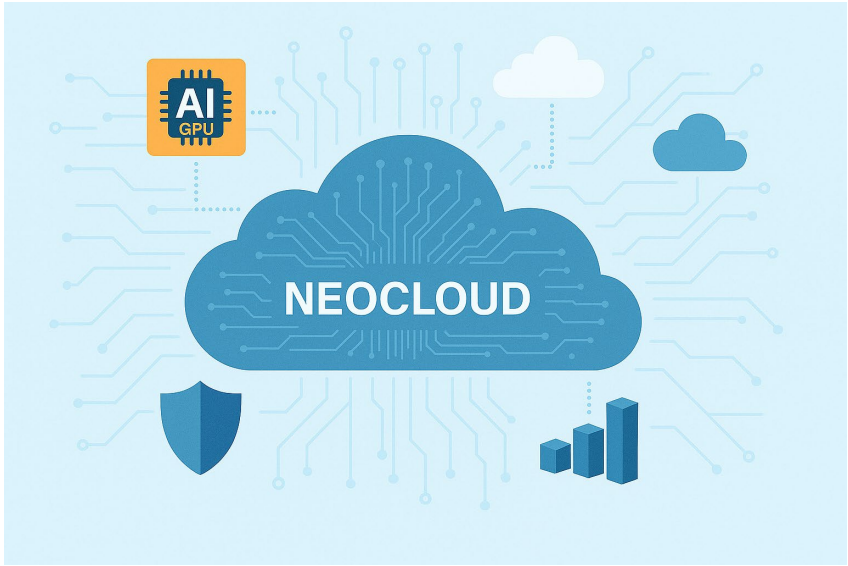
The Growing GPUaaS Market



The Rise of NeoCloud

① AI-Ready

② Distributed



② Security

④ Flexible

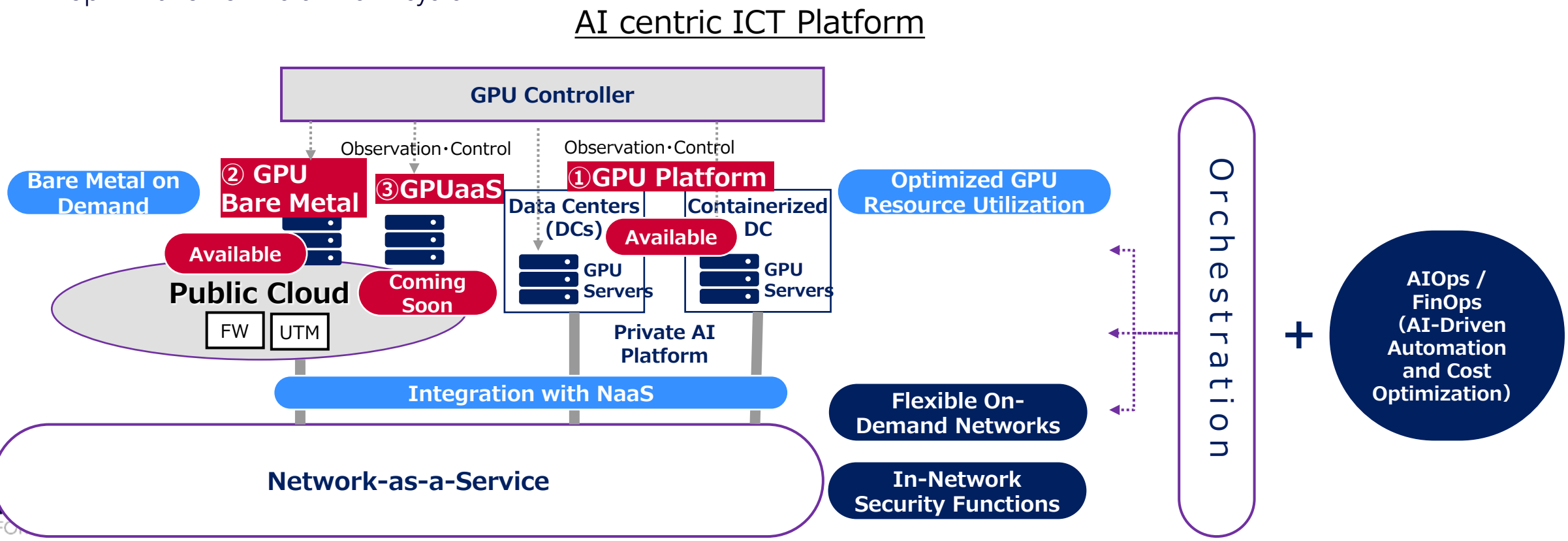
NeoCloud opens new opportunities for telecom, enabling them to leverage strengths in data centers, MEC, NaaS to create new revenue.

AMPLIFY FORMERLY MEF

GPU As A Service Market (2025 – 2030), Grandview Research.

NTT DOCOMO BUSINESS's Initiatives in GPUaaS

- **GPU Platform:** Providing dedicated GPU environments, individually designed and built for each company, on DCs and containerized DCs.
- **GPU Bare Metal:** On-demand GPU bare metal servers available on public clouds – launched in **October 2025**.
- **GPUaaS:** Providing on-demand GPU resources with fine-grained allocation, along with essential AI development tools – **to be launched soon**.
- **Integration with NaaS:** By flexibly and on-demand controlling all GPU offerings together with NaaS, achieve end-to-end optimization of the entire AI system.



Key Features of NTT DOCOMO BUSINESS's GPUaaS

01 Flexible Use of GPU Resources

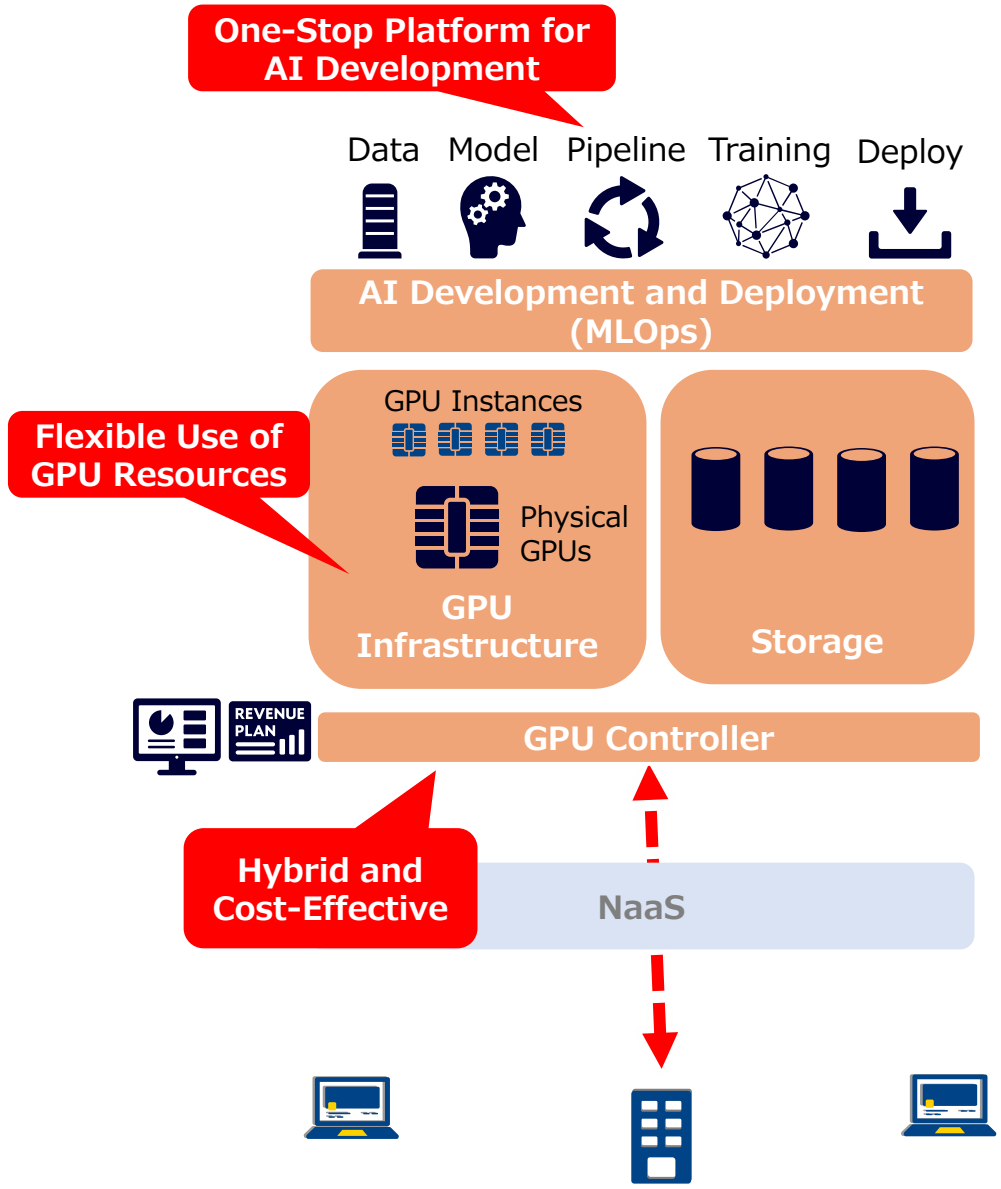
- Available **per GPU unit** with **no upfront investment**
- Usable **in fractional units** (e.g., 0.5) **per user**

02 One-Stop Platform for AI Development

- Available with essential **AI development functions**
  data, model, and pipeline management, training, and deployment
- **Available with AI inference** and AI **agent execution environments**
  featuring endpoint access tracking and control

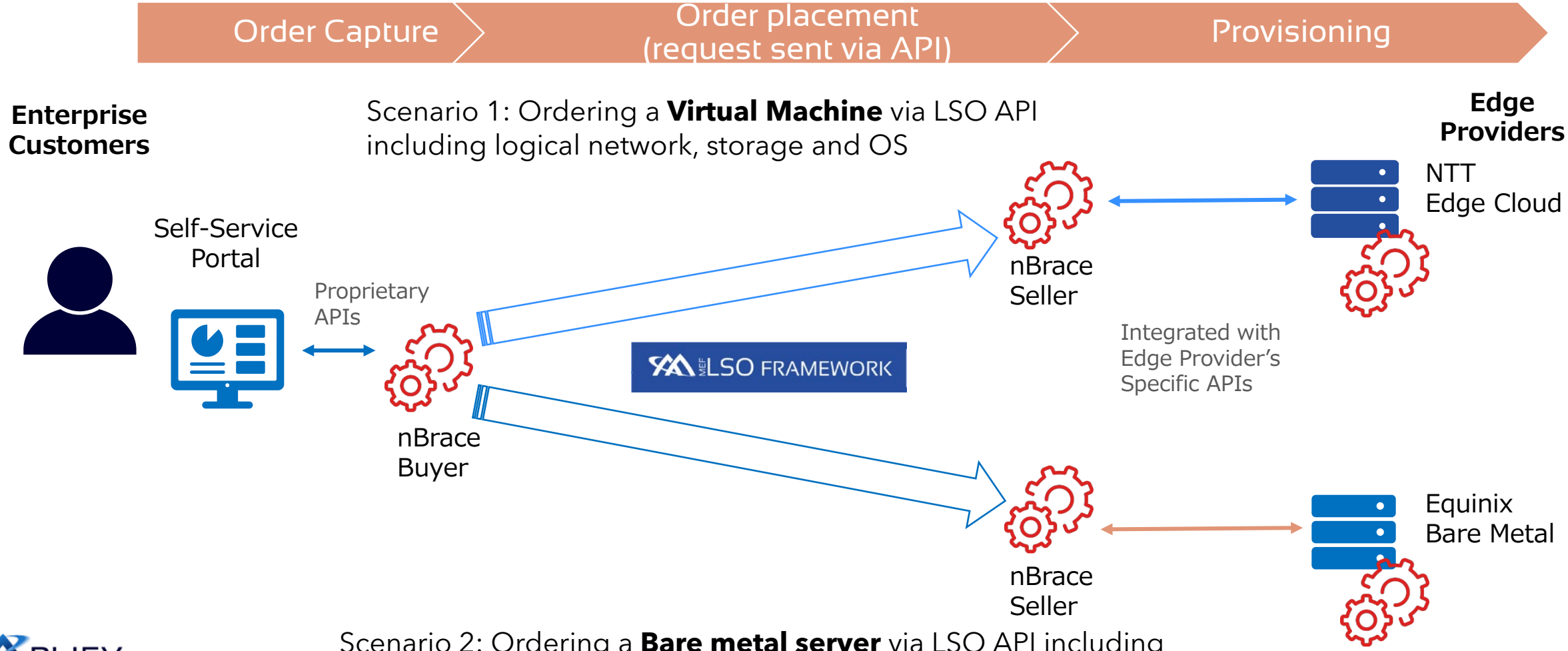
03 Hybrid and Cost-Effective

- Manageable under **a unified UI and GPU controller** across GPU Platform, GPU Bare Metal, and GPUaaS
- Available with **no data transfer fees** and **low-cost storage**



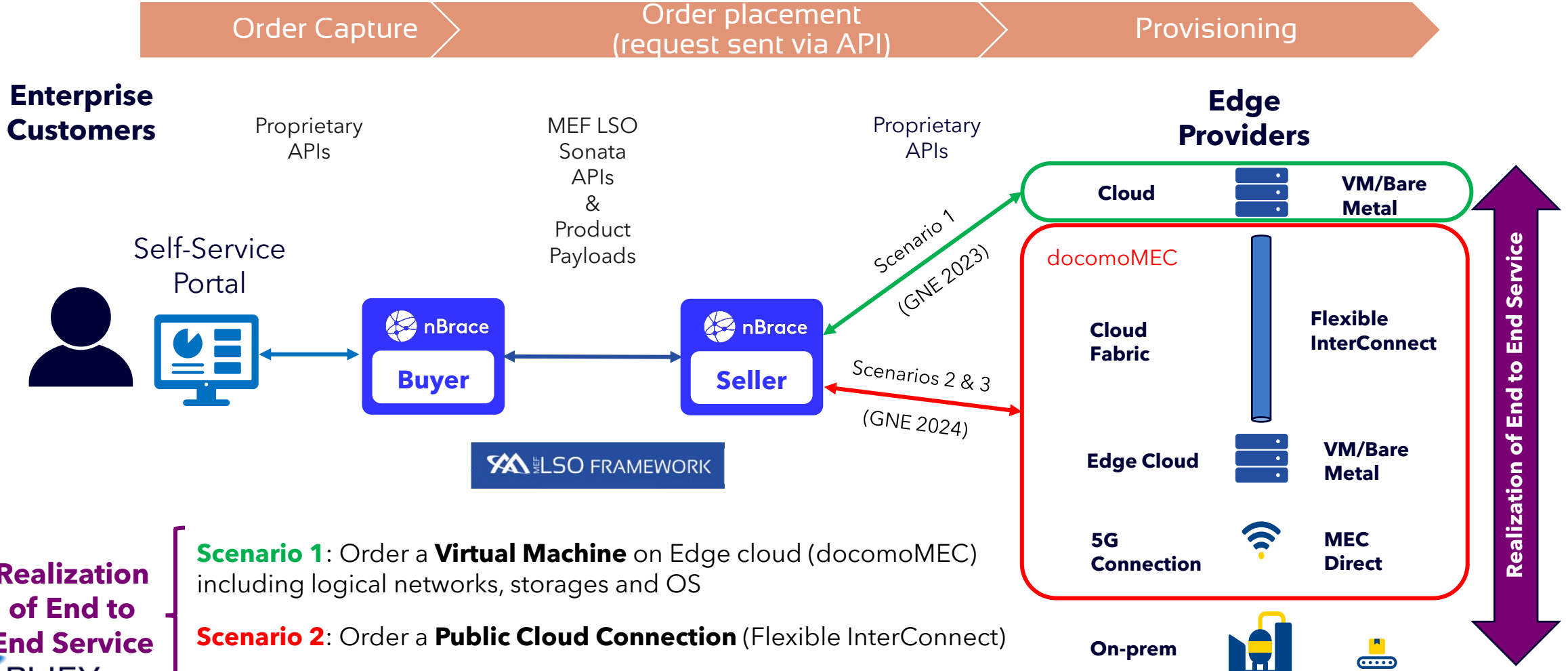
Hot Mongoose Project (2023)

- Enabled the automated buying and selling of IaaS, facilitating the growth of multi-provider telco edge and expanding the available footprint to enterprise customers through a single supplier.



Hot Mongoose Project (2024)

- Expanded from edge cloud ordering to include Cloud Fabric and 5G connectivity, enabling end-to-end service access from devices to the cloud in a single process.



Realization of End to End Service

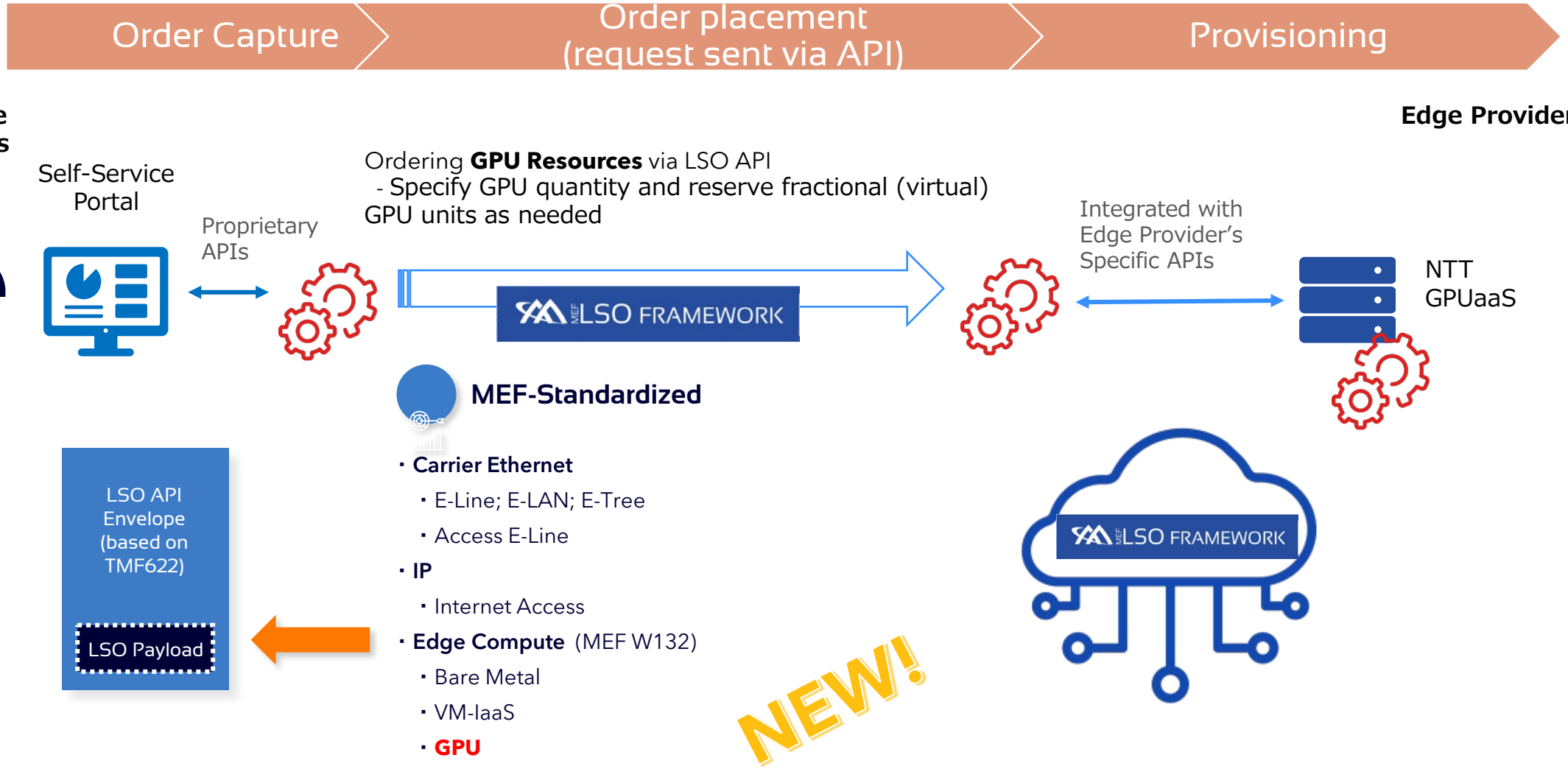
Scenario 1: Order a **Virtual Machine** on Edge cloud (docomoMEC) including logical networks, storages and OS

Scenario 2: Order a **Public Cloud Connection** (Flexible InterConnect)

Scenario 3: Order a **5G Connection with a SIM card** (docomoMEC Direct)

Hot Mongoose Project (2025~)

- Expanded the MEF LSO payload to support GPU resource operations, enabling **on-demand GPU reservation** and making **multi-provider GPUaaS** possible.

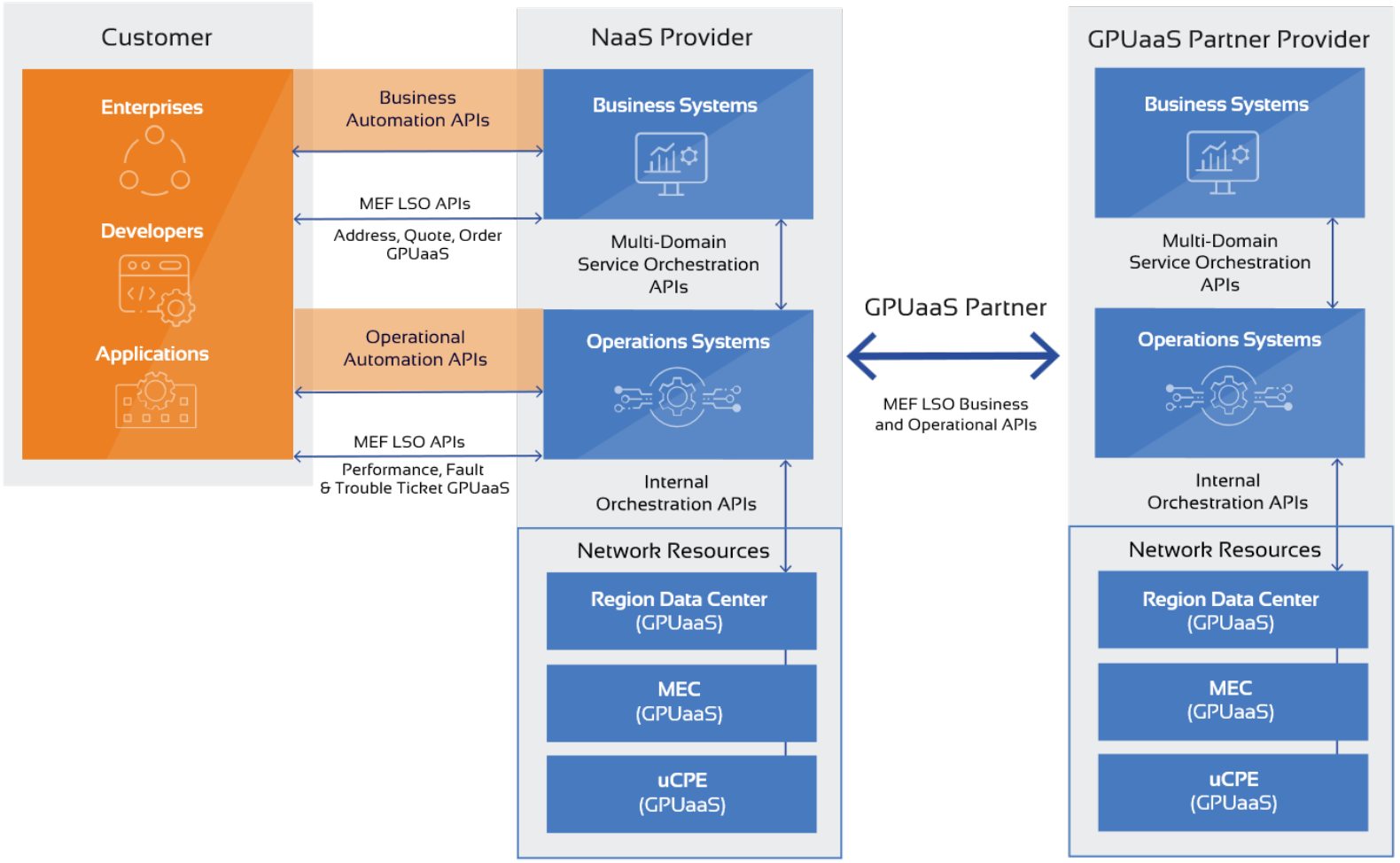


Abhishek Singhal
Industry Principal (Telecom)
Infosys Limited

GPU as a Service

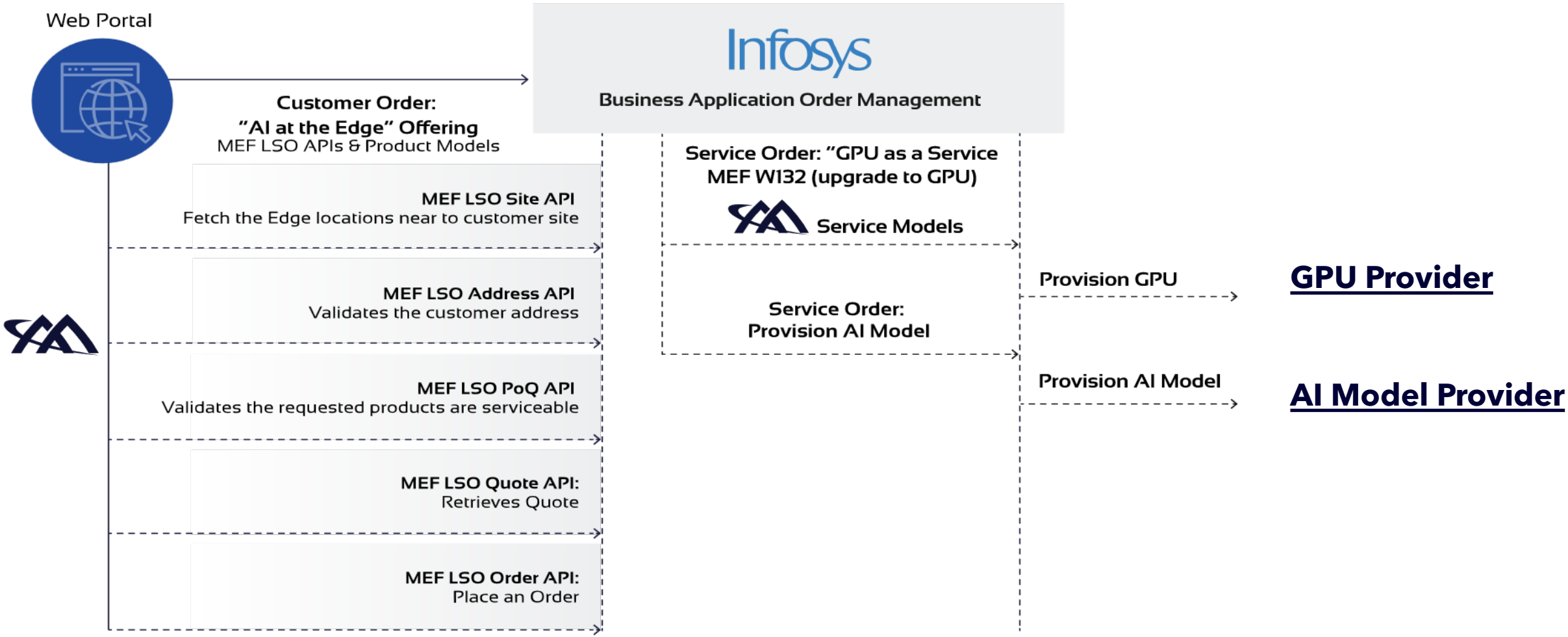
Enabling Next-Gen AI Workloads

GPUaaS

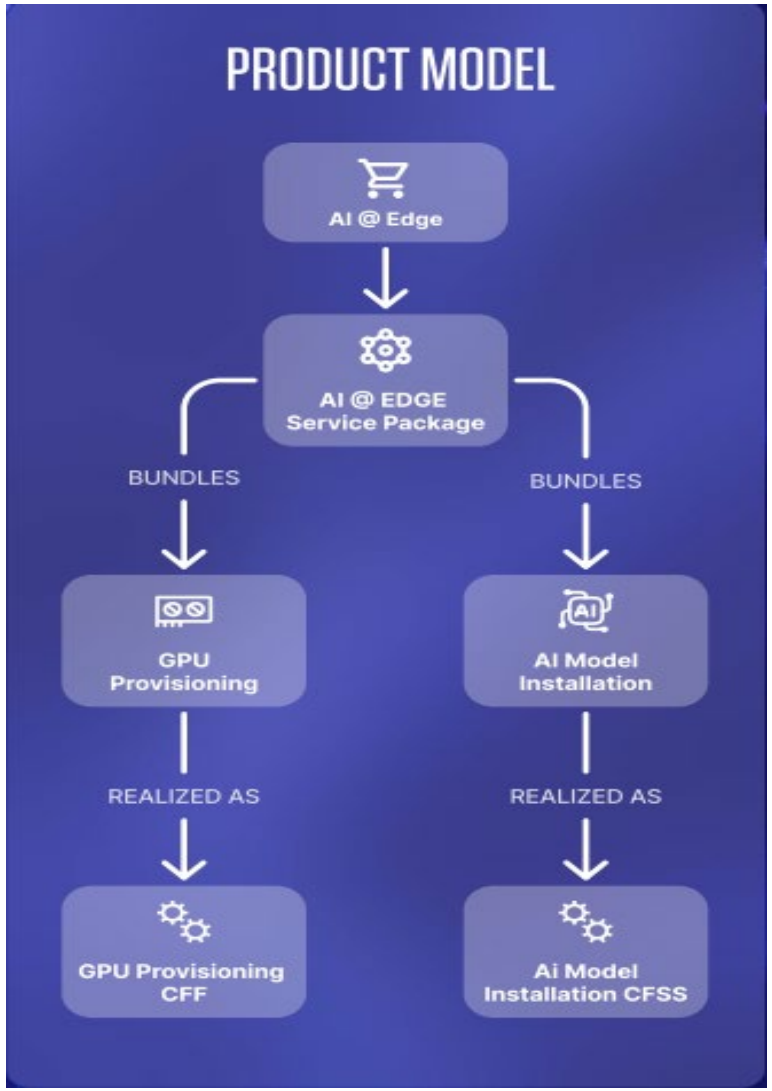


- Shared, on-demand GPU infrastructure accessible via APIs
- GPUaaS leveraging MEF LSO Business & Operational API's
- Extending MEF W132 by adding GPUaaS schema
- Accelerates time-to-value for AI deployments for enterprise customers with pre- trained models.

GPUaaS for Enterprises



Product Model



VirtualMachineWithGpuaas	
packageType	PackageType
processorType	ProcessorType

GPU	
cudaCores	integer
tensorCores	integer
performanceProfile	GPUPerformanceProfile
supportedFrameworks	GPUSupportedFrameworks[]
vram	integer
driverVersion	string

Real World Applications

- Smart Surveillance & Video Analytics for Retail
- Fleet Management with AI-based Telematics
- Automated Quality Control in Manufacturing : Computer Vision
- Traffic & Heat Mapping
- Warehouse Automation for E-commerce
- Workplace Safety Monitoring

Benefits

Telcos

- **Leverage Existing Infrastructure**: Utilize well-connected Telco exchanges as strategic edge locations for GPUs.
- **Cost-Effective**: No need for new data centers; maximize the value of existing infrastructure.
- **New Revenue Streams**: Upsell GPUaaS to SME customers, expanding service offerings and generating additional income.

Enterprises / Customers

- **Affordable Access to High-Performance Computing:** Pay-as-you-go models with no upfront hardware costs.
- **Scalability:** Easily scale GPU resources based on demand without hardware limitations.
- **Low Latency:** Edge locations ensure faster processing, providing a competitive advantage for data-intensive tasks.


GNE
Global NaaS Event