R&D phase | Research **Expected Commercialization** | FY28-

Mehr Licht!
More Photonics for Intelligence

#A05

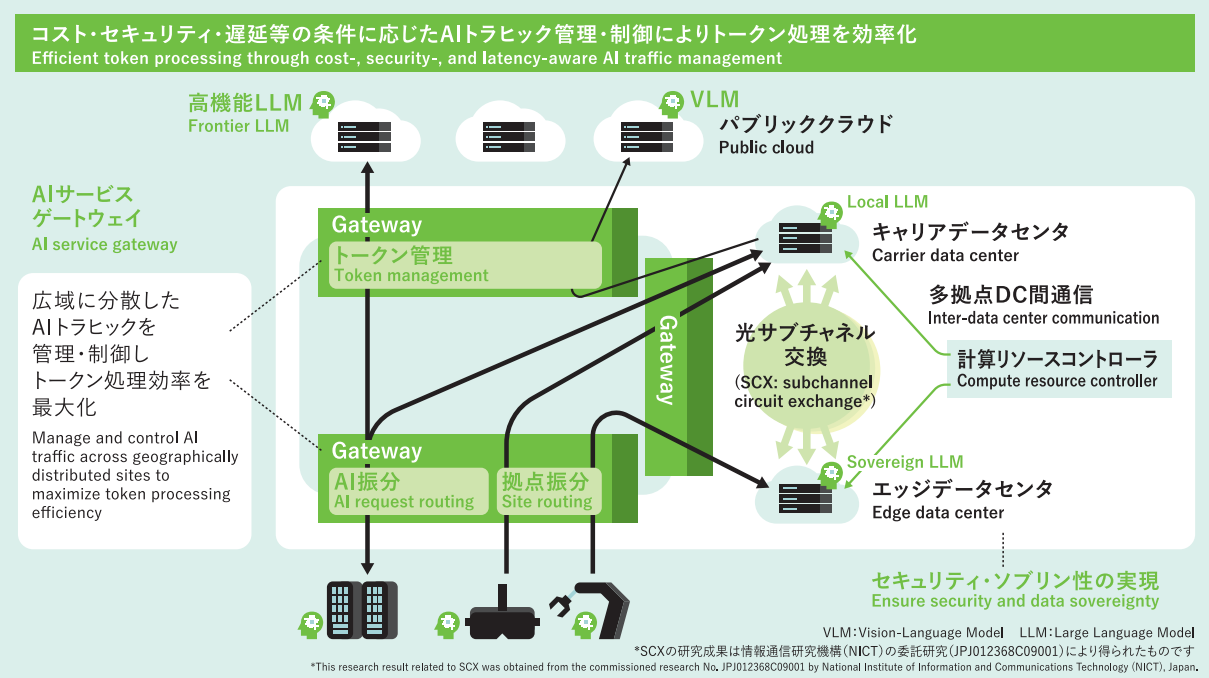
#AI & Quantum Infrastructure#Digital & AI Transformation

Traffic-Controlled Distributed AI Infra

Maximize token processing efficiency by AI traffic control

Technical Challenges & Goals

As AI demand surges, effectively using computing resources and power has become increasingly critical. By maximizing token processing efficiency, we optimize compute and energy consumption while improving response times and the overall user experience.



Key Technologies

Core Technologies

- Maximize resource utilization by dynamically allocating distributed compute based on availability
- Deterministic multi-site networking for efficient, flexible distributed computing

Key Differentiators

- Improves LLM token performance by 40% and task completion by 20% over conventional solutions
- Enabling flexible inter-DC links, assured QoS, and efficient use of optical resources

Use Cases

Information Technology (IT)

In AI-agent-driven software development, workloads are dispatched on the basis of GPU utilization, improving token processing performance while maintaining development quality.

Exhibitors

NTT, Inc. Network Service Systems Laboratories
Co-exhibitors: NTT DOCOMO, Inc., NTT DOCOMO Business, Inc.

External Partners

F5 Networks, Inc.
KDDI Corporation, Rakuten Mobile, Inc., NEC Corporation, 1FINITY Inc.

CONTACT:Network Architecture Project