

# Telcos Monetizing Digital Fabric for AI: from Cloud to Edge

Brian Washburn

Chief Analyst, Telco B2B Solutions

[askananalyst@omdia.com](mailto:askananalyst@omdia.com)

## Licensed Reprint

# Contents

- Introduction 3
- Why distributed AI orchestration matters 6
- NVIDIA AI Grid benefits and limitations for the telco edge 10
- NVIDIA AI Grid use cases: What works, what does not 14
- Pioneer service providers and their edge contexts 17
- Appendix 24

# Introduction



- **The long-awaited promise of edge services monetization could finally come to pass.**
  - In March 2026, global AI juggernaut NVIDIA released AI Grid, an orchestration platform that can organize, manage, and distribute AI models and keep them synchronized.
  - This digital fabric for an AI platform can match requests to distributed resources based on cost, performance, and workload type.



- **NVIDIA's move aims to address a vital missing piece for service providers aspiring to offer AI at the edge.**
  - The platform helps remove friction and complexity in distributed workload management as AI usage continues to ramp up.
  - Demand for latency-sensitive workloads did not materialize when communication service providers (CSPs) first started marketing multi-access edge computing (MEC) services about five years ago. AI has the potential to change the demand equation.



- **CSPs can provide a distributed fabric for AI workloads.**
  - NVIDIA's initial AI Grid partners include AT&T, Charter Communications, Comcast, and T-Mobile in the US; Indosat Ooredoo Hutchison of Indonesia; and US-based global edge specialist Akamai. Other service providers will join over time.
  - Each partner is trialing different angles, including consumer, small business, enterprise, public sector, and internal uses.



- **The edge ecosystem is growing wider and deeper.**
  - A separate, Europe-led federated telco edge alliance includes Deutsche Telekom, Orange, Telefónica, TIM, and Vodafone.
  - Initial NVIDIA AI Grid functionality showcases include Personal AI (virtual assistant); Linker Vision (smart cities); Decart.ai (real-time generative video); and Serve Robotics (local delivery). These edge use cases can benefit from, but don't depend on, AI Grid.

# Recommendations for communication service providers



**Find more compelling use cases:** CSPs cannot avoid doing their homework. They must find use cases where a distributed AI fabric is less hassle, performs better, or improves economics over existing alternatives. NVIDIA's own suggested use cases show that unique edge AI use cases remain thin on the ground.



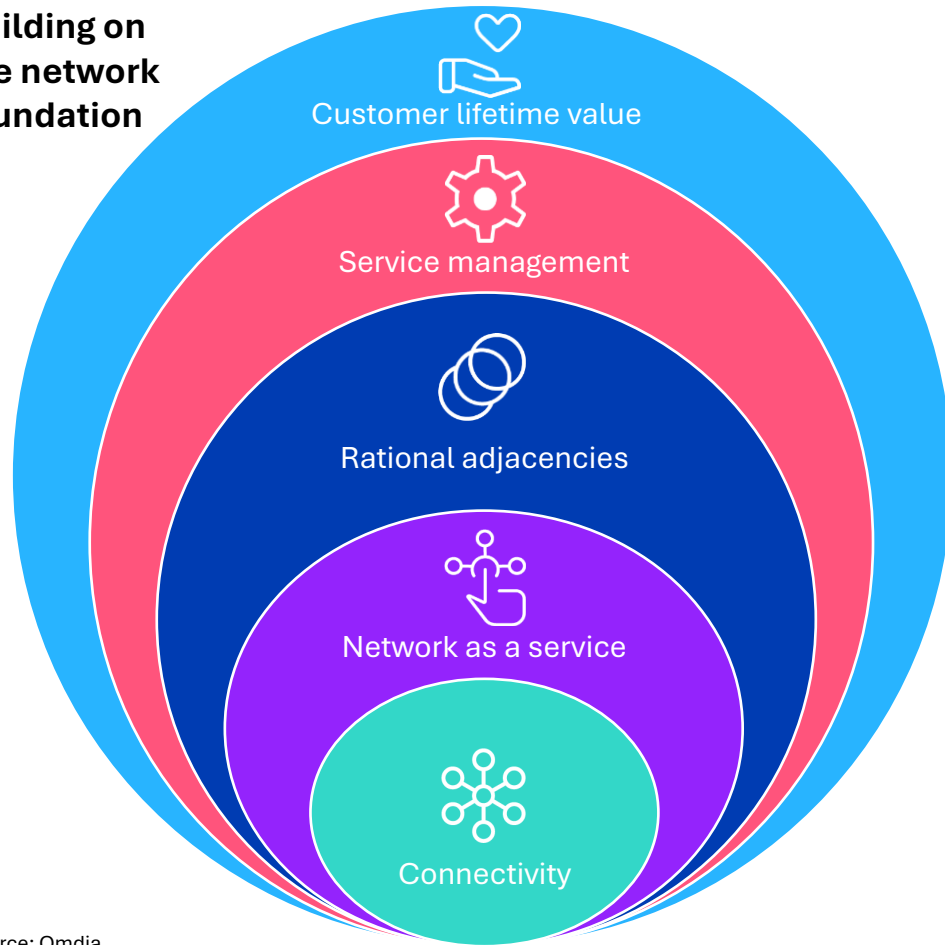
**Keep building cross-network edge alliances:** There is strength in numbers. Regional and global edge provider alliances remove friction for enterprises with far-reaching needs. But the devil remains in the details—not just operational, but also contractual gaps need filling to provide certainty to future buyers of such distributed edge AI services.



**Lean into certified security and sovereignty:** CSPs operate highly regulated infrastructure that they have invested in extensively. Certifications and competency audits are both a bar to entry and a marketing tool, not simply a cost. They are a foundation for trusted, sovereign edge services management.

# How a digital fabric for AI fits into telco B2B's trillion-dollar growth mission

Building on the network foundation



Value layering

Higher

Single view of the truth



Industry solutions, managed security



Digital Fabric for AI



Hybrid connectivity

Lower

- A digital fabric for AI represents:
  - A **foundation** underpinning service evolution beyond pure connectivity.
  - A **platform** for rapid service configuration, deployment, and change management.
  - A **cross-selling enabler** as customer needs change and new services emerge.
- Digital fabric for AI supports the goal to grow global telco B2B revenue beyond \$1tn in the next five years.

Source: Omdia

© 2026 Omdia

# Why distributed AI orchestration matters

# Why distribute AI processing: Better access for users and resources



### Real-Time Latency

Less round-trip travel time between requests and responses, necessary for true real-time interactivity.



### Efficient Localization

Local data transfer, processing and storage reduces loads placed on wide-area networks.



### Reliable Traffic Delivery

Reduced hops for local endpoints means a lower chance of congestion and dropped traffic.



### Robustness via Failover Options

Distributed processing yields other options when some network routes or AI nodes are congested.

# The components and intent of NVIDIA AI Grid

## NVIDIA's top-to-bottom platform for AI Grid

- **Policy engine:** understands intent and available distributed resources. It connects the two for the best outcome.
- **Platform layer:** offers software development kits plus load management, microservices support, and software libraries.
- **Containerization layer:** wraps workloads in Kubernetes containers, which can map across virtual machines, bare metal, and clusters.
- **Foundation software stack:** Manages compute, storage, database, and networking functions, with security, load balancing, and cluster lifecycle.
- **Hardware Stack:** consisting of select HGX and RTX series compute platforms.



## CSPs: initial, but non-exclusive targets

- **AI Grid was designed with four goals in mind:**
  - Support distributed (GPU) computing at the edge;
  - Work on top of a high-speed/low-latency fabric;
  - Offer workload routing through a control plane that is aware of/adheres to SLA requirements;
  - Support a unified deployment model (including multi-tenant AI).
- **AI Grid itself does not specify WAN performance.**
  - Service providers decide their own network performance criteria.
  - Policy engine routing requests to nodes draws on network-relevant factors such as lowest latency.

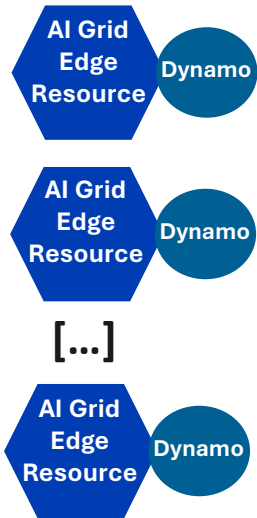
Source: Omdia

© 2026 Omdia

# Management components of NVIDIA AI Grid orchestration

## NVIDIA Dynamo

- Distributed GPU, traffic, and storage management in the data center
- Optimizes distributed inference across GPUs



## Network routing

SDN controller  
(software defined network)

- Establishes network connectivity and resourcing

## AI workload routing

AI Gateway/LLM router

- Analyzes requests and directs to the most suitable model based on complexity and context
- Optionally supports caching (responses to frequent requests)

## AI Grid orchestration

- Tracks availability of capacity & resources
- Manages resources (pods, clusters, AI models)
- Detects failures, reprovision the model elsewhere

## Policy engine

DPE (decision and policy engine)

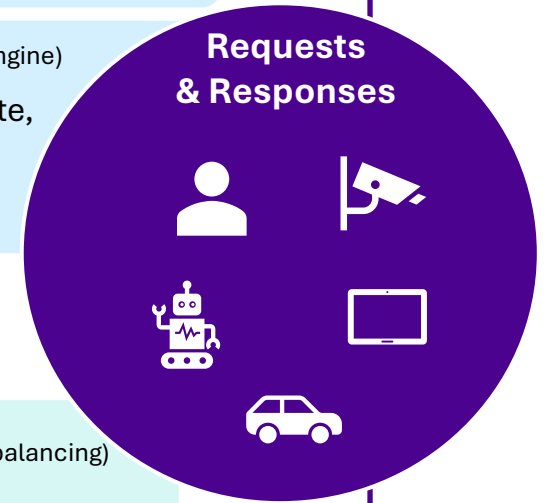
- Decides what resources to allocate, based on factors such as cost-efficiency and governance

## Load balancing

GSLB  
(global server load balancing)

- Determines where to send requests for processing based on factors such as GPU utilization

## Requests & Responses



Source: NVIDIA, Omdia

# NVIDIA AI Grid benefits and limitations for the telco edge

# The industry problem of AI resource orchestration

## Internal orchestration: An intra-domain technology challenge

- **Technology challenge:** Providers need to run distributed instances of AI models, with a dynamic understanding of available inventory and status, routing logic for workload requests, while keeping AI instances synchronized and properly calibrated.
- **Work is underway:** NVIDIA's AI Grid aims to solve technical challenges of matching requests and resources, while keeping distributed AI models synchronized and managing underlying hardware and software platforms.
- **Who is involved:** Edge providers—mainly CSPs (telcos and cablecos)—are initial pilots for AI Grid.
  - But AI Grid does not need to be exclusive to network operators. Any distributed AI hosting business could use it—hyperscalers, neoclouds connecting distributed facilities, or specialists building AI platforms.

## Multi-party orchestration: A separate, inter-domain business challenge

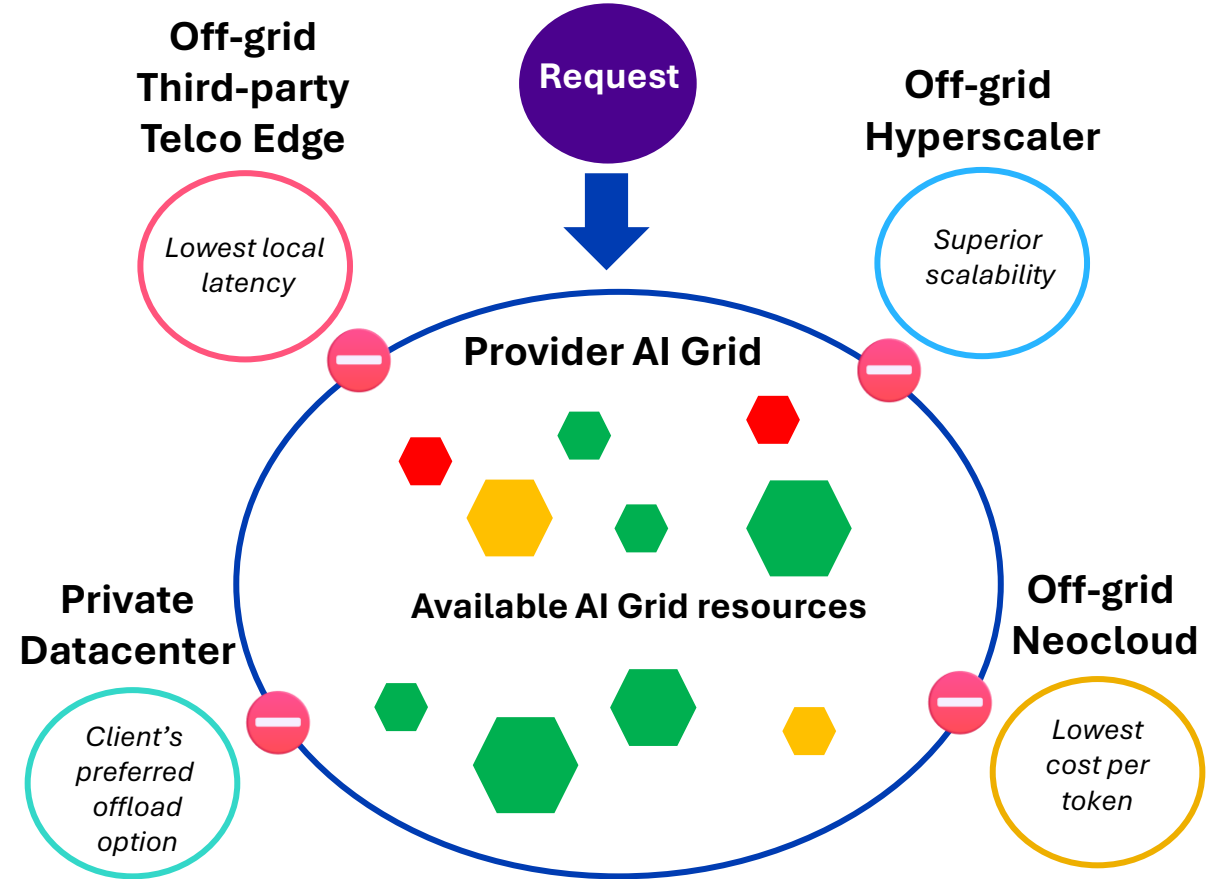
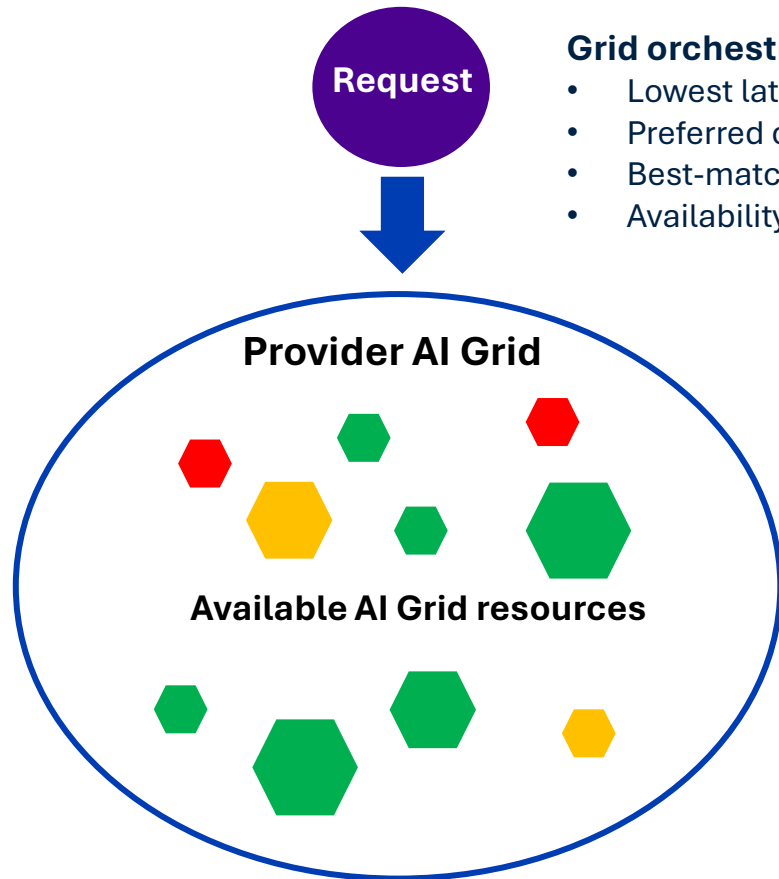
- **Business challenge:** Edge players need to agree on common processes for contracting, provisioning, and operating on each other's edge resources. This is not a vendor technology issue, but a framework and process standards issue.
- **Work is underway:** The European Cloud-Cloud-Continuum (EURO-3C) for telco edge cloud is part of the Important Project of Common European Interest on Next Generation Cloud Infrastructure and Services (IPCEI-CIS) for multi-provider edge. This is, in turn, a project of the 8ra initiative, funded by the European Union.
- **Who is involved:** Vodafone, Orange, Deutsche Telekom, Telefónica, and TIM have demonstrated a pan-European federated edge compute infrastructure.

## Demand drivers: Carriers need to meet hyperscaler and other AI supplier expectations

- **Business challenge:** Providers that want to sell AI Grid hosting need to convince hyperscalers and other AI industry players of the benefits of their distributed AI hosting/federated edge services.
- **Work is underway:** Initial AI Grid adopters are still exploring what buyer incentives will land paying customers. It's one thing to design the technology and build processes, but another to have AI Grid's route to market succeed financially against alternatives.
- **Who is involved:** Six service providers have initially showcased NVIDIA AI Grid with pilots and/or rollout commitments. These are Akamai, AT&T, Charter Communications, Comcast, Indosat Ooredoo Hutchison (IOH), and T-Mobile USA. Three more operators (SK Telecom, Softbank, Telus) are not formal AI Grid adopters, but they have teamed with NVIDIA on deeper AI efforts to date.

# NVIDIA AI Grid is designed to embed, distribute in a network...


# ...but lack of off-grid orchestration limits AI Grid's functionality



# NVIDIA AI Grid use cases: What works, what does not

# More use cases are needed for AI distribution at the edge

## Three use cases aligned to NVIDIA's AI Grid



### Gaming GPUs

NVIDIA's GeForce NOW hosted GPU service is not exactly AI. But it is a practical, in-demand service for GPU processing power, which converts consumers to \$10 and \$20 monthly subscription plans.

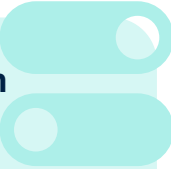
AI Grid can bring GPUs closer to gamers: The network operator can win the business: It reduces traffic on its networks while providing gamers with low-latency experiences.



### Sovereign mission

Nations that aim to offer AI as a public resource, which are also concerned about AI sovereignty, can turn to domestic telcos to guarantee security, control, and localization of the platform and its data.

AI Grid can serve each municipality its own local AI resources, giving local governments and communities buy-in while the network operator administers the national infrastructure.




### Machine monitoring/automation

Consumer/mass-market physical AI and XR experiences are still some years off. But there is a fast-evolving AI role in industrial monitoring and control automation that benefits from nearby processing and storage.

There are flexibility and performance benefits to sending industrial images and video into AI hosted at the edge vs. onsite or centralized alternatives.


# AI Grid's initial use cases fail to differentiate the edge service provider

- Efficient localization, reliable performance and robust failover are potential benefits of putting AI at the edge nearer users.
- But reasonable alternatives already exist for each of the suggested use cases for AI Grid. Centralized AI is adequate for some; other uses are strictly local and do not need distributed AI's benefits.

**Scenario: Personal assistant** 


✓ **Pro:** Low enough latency to deliver positive user experiences.

✗ **Con:** Natural human conversation has loose latency (e.g., <300–500ms). The onus is on AI model efficiency and assigned processing power. Subscribers may travel outside locales and expect similar performance. Performance improves more from using smaller language models.

**Scenario: Smart city** 


✓ **Pro:** Sensible to hub/store all information feeds and content in-metro. Critical information: safety/emergencies, traffic, and hazards.

✗ **Con:** Localized AI does not benefit from AI Grid's load distribution. Critical smart cities information is rarely latency sensitive. Strong citizen pushback to cameras in certain public spaces.

**Scenario: Local delivery** 

✓ **Pro:** Robots and drones need to make real-time dynamic decisions. They need to avoid potentially disastrous failures (collisions). Limited on-board AI processing due to cost and weight.

✗ **Con:** Localized AI does not benefit from AI Grid's distribution system. Sidewalk delivery is limited to low speeds (3–5mph), which tolerates latency. Robot vehicle delivery faces other challenges.

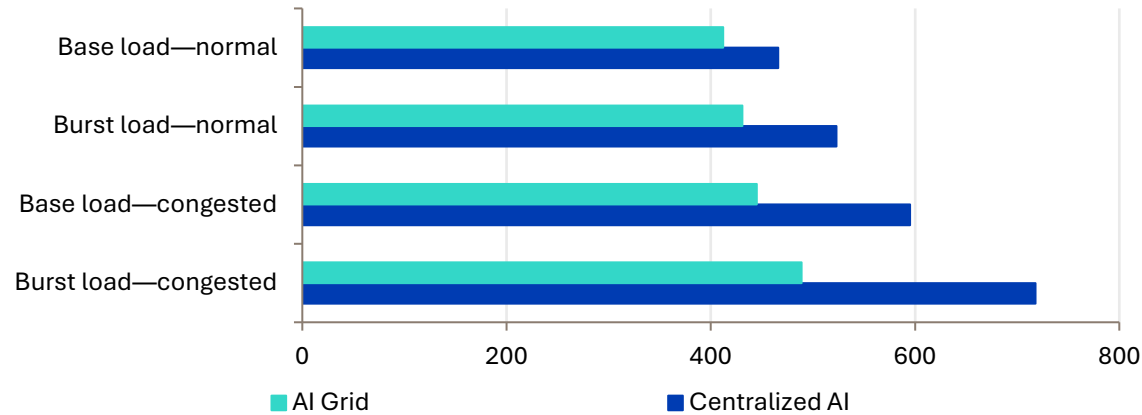
**Scenario: Generative livestream** 

✓ **Pro:** Overlay video supports real-time interaction. Low local latency synchronizes users' lips, movement, and expressions.

✗ **Con:** Live streaming is one-way and does not change viewers' experiences. There is no need for real-time latency as viewers respond by text. Frequent VTubers will likely prefer on-device AI video processing over hosting services.

# Measurable improvements claimed for AI Grid

Voice response latency, AI grid vs. centralized AI

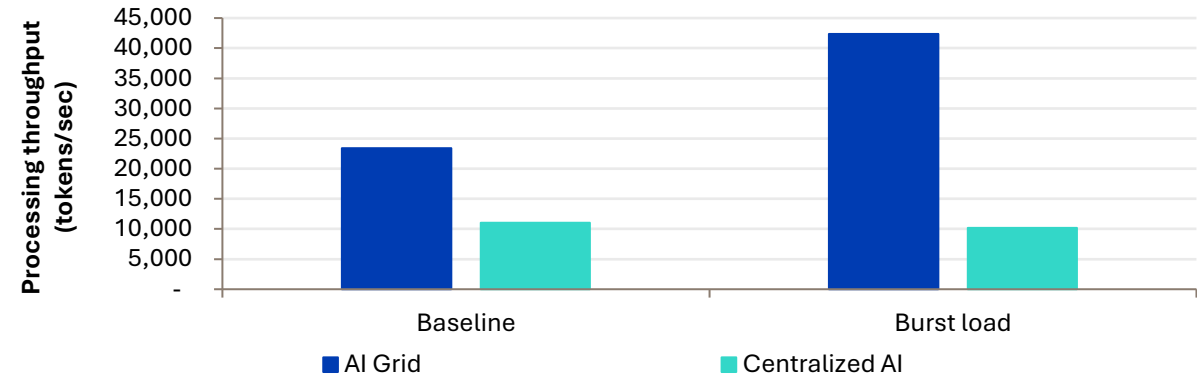


Source: NVIDIA, Omdia

© 2026 Omdia

- For a voice interaction scenario, AI Grid demonstrated baseline 54ms lower end-to-end latency (13% improvement) over centralized AI.
- Under various loading scenarios, AI Grid's more robust distribution kept latency lower. Examples showed centralized AI latency was 92, 150, and 229ms slower than AI Grid.

Processing load, AI Grid vs. centralized AI



Source: NVIDIA, Omdia

© 2026 Omdia

- At baseline, AI Grid handled 2.1x the load of centralized AI when limited to strict end-to-end applications latency.
- Under a burst (high concurrent user) scenario, AI Grid could handle 4.1x more than centralized AI under end-to-end applications' latency restrictions. AI Grid's lower network latency freed up time for more compute queuing and processing.

# Pioneer service providers and their edge contexts

# AI Grid ambitions: Akamai's global AI enablement of its edge



## Purpose

- Unlike its peers, Akamai has already committed to large-scale AI Grid adoption. The company intends to roll out “thousands” of GPU servers globally.
- The stated intent is to assist enterprises interested in running low-latency agentic and physical AI.
- Given the size of its distributed infrastructure commitment, Akamai highlights the importance of AI Grid's intelligent orchestration to leverage an architecture with thousands of nodes.



## AI Grid pilots

- The company has not listed active pilots with third parties. Akamai had previously noted goals to support distributed agents with faster responsiveness, support for latency-sensitive physical AI, and hosting AI models at optimal locations for clients.
- Akamai's global distributed network has more than 4,400 locations worldwide.



## Prior activity

- In October 2025, Akamai announced its AI Inference Cloud, a service that uses NVIDIA infrastructure to bring inferencing to the edge, with 20 target deployment locations worldwide.
- Akamai's role as an edge player goes back many years, given the company's roots in content delivery networks (CDN), traffic optimization, and a wide collection of distributed services in addition to content.

# AI Grid ambitions: AT&T's AI IoT enablement with Cisco



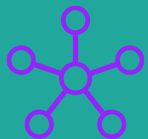
## Purpose

- AT&T is implementing NVIDIA AI Grid's reference architecture through partner Cisco's offer, branded "AI Grid with NVIDIA".
- The focus combines many elements appealing to AT&T: Its nationwide wireless network and IoT connectivity; its past investments in edge; and mature partner suppliers, plus a larger ecosystem.



## AI Grid pilots

- AT&T demonstrated the joint technology for public safety with AI functionality from Linker Vision, connected by AT&T IoT Video Intelligence (an upstream, streamlined video aggregation service).
- AT&T has a pilot with industrial enterprise TanMar (Eunice, LA, US) that includes physical surveillance/security and license plate reading.



## Prior activity

- Prior edge activity includes a 2021 deal to make AT&T Network Edge available with Google Cloud.
- A Microsoft Azure agreement in March 2026 makes available turnkey intelligent edge solutions, branded as AT&T Connected Spaces for Enterprise, for industries including retail, restaurants, medical, and hospitality.
- AT&T's edge activity dates back to the 2010s. Over the years, AT&T has made cloud/edge deals, including Microsoft and IBM, and established a partnership with Tech Mahindra.

# AI Grid ambitions: Comcast Business' edge trials



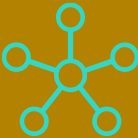
### Purpose

- Trial goals will validate latency improvement, power/cost efficiencies, resiliency, scaling, and quality of user experience.
- Comcast is increasingly drawn to value-added services that can better monetize its extensive access networks and customer relationships. AI Grid potentially provides value-added subscription opportunities.



### AI Grid pilots

- Comcast's field trial investigates three functions delivered from the network edge:
  - Customized advertising, with micro-targeting down to the individual household (partner: Decart.ai).
  - A small business personal assistant that answers incoming questions and manages appointments (partner: Personal AI).
  - Low-latency gaming through the NVIDIA GeForce Now service, with GPUs located closer to gamers who subscribe to these hosted GPU services.



### Prior activity

- Comcast has been active in edge for years. Its activities include compute nodes embedded in its access network.
- Comcast previously launched DOCSIS for gaming with Valve Steam and NVIDIA GeForce Now. It has shown video/collaboration/immersion with Apple TV/Apple Vision Pro, and apps using Meta mixed-reality headsets.
- Comcast's partners include Vapor IO, which was to explore opportunities such as hosted computer vision, video security, private 5G, and edge cloud services.

# AI Grid ambitions: IOH distributes public AI resources in Indonesia



## Purpose

- Indosat Ooredoo Hutchison (IOH) plays a part in Indonesia's national social mission to democratize accessibility to AI.
- The provider is using AI Grid to support distributing its existing national AI ecosystem, to bring compute power out of Indonesia's largest cities and closer to domestic communities, developers, and businesses.



## AI Grid pilots

- Distribution of Indosat's existing AI models and national infrastructure—where NVIDIA has played a role from the outset. AI Grid helps back the provider's civic Sahabat AI open platform, which is accessible to citizens through Android and iOS applications.



## Prior activity

- IOH forged a memorandum of understanding with NVIDIA in 2024 to supply its subsidiary Lintasarta with AI-capable GPU platforms, powering its GPU Merdeka sovereign AI cloud and Deka GPU services.
- In 2025, Indosat partnered with Cisco and NVIDIA to establish an AI center of excellence, which included AI-powered security services. It also partnered with Nokia and NVIDIA to explore AI-RAN and AI grid uses.
- In 2025, Indosat first launched its domestic/sovereign Sahabat AI LLM, supporting local languages.

# AI Grid ambitions: Spectrum Business' CGI rendering demo



## Purpose

- Spectrum Business is a major national US fiber asset holder. The provider connects to more than 300,000 buildings and attaches to more than 400 data centers, plus hyperscalers and cloud providers.
- The provider has more than 1,000 edge locations, enabling sub-10ms latency, plus hundreds of Megawatts of spare electrical load. This makes AI Grid the foundation for a planned customer/partner ecosystem.



## AI Grid pilots

- Spectrum Business initially demonstrated an AI Grid use case to render CGI for film studios. The pilot is served from Spectrum's edge, paired with 100 Gbps connectivity for rapid professional video production.
- The pilot showcases infrastructure that handles a range of real-time/latency-sensitive enterprise workloads with additional partners as the provider builds out its ecosystem.
- Other functionality under consideration includes GPUs for gaming, conversational AI, and physical AI.



## Prior activity

- The provider formally introduced its 1,000-site Edge Compute Infrastructure (ECI) as part of its AI Grid announcement, but these edge resources pre-date this latest partner announcement with NVIDIA.
- Spectrum Business has a history of funding large-scale fiber buildout and optical platform modernization. The provider has had network-wide 100 Gbps service capability since the end of 2022.

# AI Grid ambitions: T-Mobile US's edge pilot programs with Nokia



### Purpose

- T-Mobile is working with Nokia and aims to build a partner ecosystem to support physical AI via distributed edge AI.
- T-Mobile intends to leverage AI-RAN infrastructure, and it is working with NVIDIA partners in physical AI. The physical AI/AI-RAN combination aims to optimize 5G connectivity with edge AI workloads.



### AI Grid pilots

- Smart cities, which include optimizing traffic flows (partners: Linker Vision, Inchor, Voxelmaps).
- Utility reading, transmission line monitoring, and alerting (partners: Levatas, Skydio). Facilities management beyond simple sensors (partner: Vaidio).
- Industrial worksite safety (partner: Fogsphere).



### Prior activity

- In October 2025, T-Mobile launched Edge Control, a local network breakout service that can route traffic directly to on-premises edge platforms and applications.
- T-Mobile's prior hosted edge activities include a 2021 deal with Lumen for its Edge Computing Platform.

# Appendix

# Appendix

## Methodology

This report is based on attendance and presentations at NVIDIA’s annual GTC event, plus announcements at Mobile World Congress (all held March 2026). Additional information from NVIDIA documentation and discussions with provider executives participating in AI Grid; and from prior Omdia research on telco edge, cloud/network orchestration, and edge requirements for AI and conventional enterprise applications.

## Further reading

[“European telcos unite for edge, but hyperscalers need reasons to play”](#) (February 2026)

[AI Network Traffic Report: 2026 Update Analysis](#) (February 2026)

[AI Network Traffic Forecast: 2023–35](#) (February 2026)

[Telco B2B AI Monetization Index FY25](#) (February 2026)

[Telco Edge Computing Forecasts 2025](#) (December 2025)

[“T-Mobile Edge Control is a simple, effective solution to enable edge computing on 5G”](#) (November 2025)

[“Service Provider Spotlight: Comcast Business ramps up AI and adds dedicated internet to broadband infrastructure”](#) (May 2025)

[Billions at Stake: New Telco Revenue Opportunities in the Growing AI Services Space, 2022–30](#) (January 2025)

[Service Provider Spotlight: Spectrum Enterprise grows fiber-fed spread with integrated packages and a focus on ease of doing business](#) (July 2024)

[AI at the Network Edge: Implications on Traffic and Network Evolution](#) (June 2024)

## Author

Brian Washburn, Chief Analyst, Telco B2B

[askananalyst@omdia.com](mailto:askananalyst@omdia.com)

# Appendix

## **Omdia Consulting**

We hope that this analysis will help you make informed and imaginative business decisions. If you have further requirements, Omdia's consulting team may be able to help you. For more information about Omdia's consulting capabilities, please contact us directly at [consulting@omdia.com](mailto:consulting@omdia.com).

## **Citation Policy**

Request external citation and usage of Omdia research and data via [citations@omdia.com](mailto:citations@omdia.com).

## Disclaimer

The Omdia research, data and information referenced herein (the “Omdia Materials”) are the copyrighted property of TechTarget, Inc. and its subsidiaries or affiliates (together “Informa TechTarget”) or its third party data providers and represent data, research, opinions, or viewpoints published by Informa TechTarget, and are not representations of fact.

The Omdia Materials reflect information and opinions from the original publication date and not from the date of this document. The information and opinions expressed in the Omdia Materials are subject to change without notice and Informa TechTarget does not have any duty or responsibility to update the Omdia Materials or this publication as a result.

Omdia Materials are delivered on an “as-is” and “as-available” basis. No representation or warranty, express or implied, is made as to the fairness, accuracy, completeness, or correctness of the information, opinions, and conclusions contained in Omdia Materials.

To the maximum extent permitted by law, Informa TechTarget and its affiliates, officers, directors, employees, agents, and third party data providers disclaim any liability (including, without limitation, any liability arising from fault or negligence) as to the accuracy or completeness or use of the Omdia Materials. Informa TechTarget will not, under any circumstance whatsoever, be liable for any trading, investment, commercial, or other decisions based on or made in reliance of the Omdia Materials.

## Get in touch

Americas

[customersuccess@omdia.com](mailto:customersuccess@omdia.com)

08:00 – 18:00 GMT -5

Europe, Middle East & Africa

[customersuccess@omdia.com](mailto:customersuccess@omdia.com)

8:00 – 18:00 GMT

Asia Pacific

[customersuccess@omdia.com](mailto:customersuccess@omdia.com)

08:00 – 18:00 GMT + 8