Cognizant® Managed AI Routing

Governed AI access across models, providers and environments.

Connect once. Govern consistently.

Enterprise teams need the freedom to choose the right AI model for each use case without creating fragmented security, cost and operational controls. Cognizant® Managed AI Routing is a managed, policy-driven routing service between enterprise applications and approved AI services. Integrated model, agent and tool traffic can be routed, governed, metered and monitored through one Cognizant managed service, giving CIO, AI platform, security and FinOps leaders a shared point of control.

- Connect applications through one consistent interface as models and providers evolve
- Apply security, privacy and usage policies across supported AI traffic
- Attribute AI consumption by team, application and use case

AI is spreading faster than it can be governed

Enterprise AI is inherently distributed. Models, data and applications may span clouds, private infrastructure, edge locations and jurisdictions. As adoption grows one application at a time, teams often select different providers, build separate integrations and implement their own controls for access, fallback and monitoring. The result can be duplicated engineering, limited cost visibility, inconsistent policy enforcement and unnecessary distance between AI workloads and the data they use. As agents begin to use tools and act across enterprise systems, this fragmentation increases operational and governance risk.

Cognizant® Managed AI Routing addresses this challenge at the request layer, providing a common routing point, policy framework and usage record for supported AI interactions.

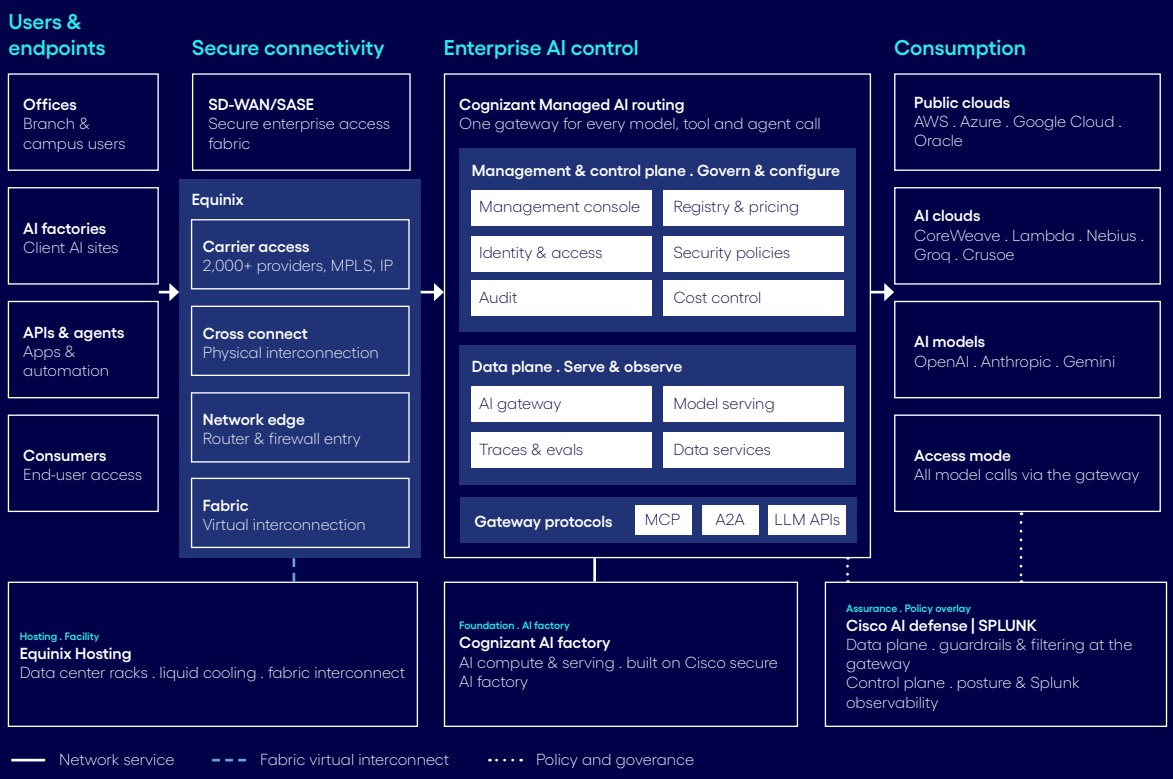
Core service capabilities

Capability	What it delivers
Policy-driven routing	Routes requests to approved models using configured policies and available signals for cost, latency, quality and data residency
Central guardrails	Applies layered access, prompt, response and data-handling policies across supported model and agent interactions. Security teams set enterprise-wide policy, and application owners add use-case rules
Usage metering	Captures token and request usage by team, application and use case to support FinOps, showback and chargeback processes
Resilience	Supports policy-controlled failover to validated alternatives when a model or endpoint is unavailable or degraded
Observability	Provides tracing, logging and configurable evaluation, with integration into enterprise monitoring and MLOps platforms
Model portfolio management	Supports model onboarding, comparative evaluation and controlled migration as enterprise requirements change

How it works

Cognizant® Managed AI Routing sits between enterprise applications and approved model endpoints. Applications connect through a consistent, OpenAI-compatible interface. Each supported request follows a governed path that can authenticate access, apply inbound and outbound guardrails, select an eligible endpoint, record usage and send telemetry to configured monitoring platforms. The service can support public, private and self-hosted models across cloud, data center, colocation and edge environments.

One control plane for every model



- **Federated deployment:** routing instances can run in the cloud and in partner colocation facilities such as Equinix, with routing and policy kept consistent across locations
- **Private connectivity:** traffic to supported model providers can use private interconnection instead of routing back through a cloud region
- **Regional control:** inference and telemetry can be kept in approved regions to support data residency objectives
- **Client-defined data handling:** prompt retention, key management and identity integration are configured to each client's security and compliance requirements

Layer	What sits there	Role
Enterprise applications	Agents, copilots and AI-enabled business applications	Connect through a consistent, governed interface
Managed routing service	Routing, policy and metering components operated by Cognizant	Applies routing, guardrails, metering, observability and failover policies
Model endpoints	Approved public, private and self-hosted models	Serve approved requests under agreed terms

Service architecture layers.

This separation helps organizations add approved providers, update policies and transition between validated models without redesigning every connected application. It creates a more consistent operating model while preserving deployment and provider choice.

From first assessment to everyday operations

Cognizant helps organizations establish and operate the service through four structured phases, tailored to the agreed scope and target architecture.

Phase	What happens	Typical outcome
Assess	Map current model usage, integration patterns, spend visibility and policy gaps	Routing readiness assessment
Design	Define the target architecture, provider portfolio, routing policies, controls and operating model	Reference architecture and implementation roadmap
Deploy	Implement the managed routing service, integrate priority applications and activate agreed policies and telemetry	Operational service for the agreed initial scope
Operate	Monitor service health and usage, evaluate model options, and maintain policies under the agreed service model	Ongoing service reporting and optimization

Engagement phases.

An open partner ecosystem

Cognizant® Managed AI Routing is built to work with an ecosystem of infrastructure, connectivity, security, observability and model partners, so clients can align the service to their existing technology choices. Current partners include Equinix and Cisco. The solution was previewed at Equinix Horizon on September 2, 2026, built on the Equinix Distributed AI Hub framework, with Cisco Secure AI capabilities supporting security and observability for the demonstrated use case.

Organization	Role in the demonstrated solution
Cognizant	Architecture, integration, policy design, model routing and managed operations
Equinix	Distributed digital infrastructure, AI-ready colocation and private interconnection through the Equinix Distributed AI Hub framework, helping place AI services closer to users and data
Cisco	Security and observability capabilities for AI applications, models and supporting infrastructure

Current ecosystem partners, as demonstrated at Equinix Horizon.

Why Cognizant

Cognizant® Managed AI Routing is operated end to end by Cognizant, from assessment and design to integration and day-to-day operations. We work across an ecosystem of partners, currently including Equinix for infrastructure and Cisco for security and observability, so clients do not have to integrate and run these layers themselves. We also help clients define an approved model portfolio and tie AI usage to the teams and budgets that own it, while keeping choice across supported models, providers and deployment locations.

- One accountable operator across routing, security, observability and infrastructure
- AI usage tied to applications, teams and cost owners to support FinOps
- Model evaluations, policy updates and controlled migrations managed through a defined operating cadence

Frequently asked questions

How is this different from an AI gateway product?

Cognizant® Managed AI Routing is a managed service, not just software. Cognizant designs, integrates and operates routing, guardrails, metering and observability with ecosystem partners, so clients do not have to assemble and run these layers themselves.

Do we need to change our applications?

Applications connect through a consistent, OpenAI compatible interface. Priority applications are connected during the Deploy phase, and the effort depends on how each application uses AI today.

Which models and providers are supported?

The service can support approved public, private and self-hosted models across cloud, data center, colocation and edge environments. The approved model portfolio is defined with each client during the Design phase.

Are we tied to specific partners?

The service is built to work with an ecosystem of infrastructure, connectivity, security, observability and model partners, so it can align with each client's existing technology choices. Equinix and Cisco are current partners.

How is our data handled?

Prompt retention, key management and identity integration are configured to each client's security and compliance requirements. Inference and telemetry can be kept in approved regions to support data residency objectives.

How do we get started?

Most engagements begin with a routing readiness assessment that maps current model usage, integration patterns, spend visibility and policy gaps, and sets priorities for the Design phase.

Let's talk

Start with a clear view of where AI traffic runs today. Cognizant can help identify priority applications, define the policies they require and show how Cognizant® Managed AI Routing can improve control, visibility and flexibility. To arrange a routing readiness assessment, contact the Cognizant AI Factory team at AIFactoryTeam@cognizant.com.



cognizant

Cognizant (Nasdaq: CTSH) is an AI Builder and technology services provider, bridging the gap between AI investment and enterprise value by building full-stack AI solutions for our clients. Our deep industry, process and engineering expertise enables us to build an organization's unique context into technology systems that amplify human potential, drive tangible outcomes and keep global enterprises ahead in a fast-changing world. See how at cognizant.ai or @cognizant.

World Headquarters

300 Frank W. Burr Blvd.
Suite 36, 6th Floor
Teaneck, NJ 07666 USA
Phone: +1 201 801 0233

European Headquarters

280 Bishopsgate
London
EC2M 4AG
England Tel: +44 (01) 020
7297 7600

India Corporate Office

Siruseri-Software Technology Park of India (STPI)
SDB Block—Ground Floor North Wing
Plot No H4, SIPCOT IT Park
Chengalpattu District
Chennai 603103, Tamil Nadu
Tel: 1800 208 6999

APAC Headquarters

1 Fusionopolis Link,
Level 5 NEXUS@One-North,
North Tower Singapore 138542
Phone: + 65 6812 4000

© Copyright 2025–2027, Cognizant. All rights reserved. No part of this document may be reproduced, stored in a retrieval system, transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the express written permission of Cognizant. The information contained herein is subject to change without notice. All other trademarks mentioned herein are the property of their respective owners.

WF 5580352