Commissioned by

OMDIA

CISCO

Contents

Published February 2025

Introduction

Ushering in a new era of IP networks in an AI-heavy future

With the recent rise of generative AI and its evolution to agentic AI, governments and businesses worldwide are crafting plans to leverage these technologies to foster digitalization and increase efficiencies. Generative AI is an essential driver of the AI era because of its impact on content generation and knowledge management among other things. Agentic AI takes things further by handling more complex tasks and autonomously making decisions that adapt to changing conditions. The market for generative and agentic AI is set for rapid expansion. However, these technologies still face challenges around reliability, transparency, and in some cases, regulatory compliance.

From a network transport perspective, one critical question is whether the existing IP transport infrastructure of service providers (including telcos, cloud, and enterprises) is ready to meet the needs of gigantic volumes of AI traffic without load imbalances and performance constraints.

The sudden growth in AI's popularity has ignited research among service providers into the modernization of their end-to-end IP transport networks. These must ensure secure, manageable, resilient, high-bandwidth, ultra-reliable, high-throughput, and energy-efficient sustainable connectivity for their business and residential customers.

This paper examines the impact of AI workloads and suggests how IP transport networks must evolve to support them. It explores the following key questions:

- How much AI traffic will service provider networks have to carry?
- What network resources will this AI traffic consume?
- Where should AI traffic be processed to achieve low latency?
- Which areas of the IP network need to be modernized first?
- How should IP networks be optimized to maximize revenue and minimize total cost of ownership?
- How can AI enable network automation and increase operational efficiency?
- How can network security be maintained in the AI and quantum eras?

The paper provides upgrade guidelines and recommends a technology tool kit for modernizing IP metro and core networks.

Published February 2025

Navigating challenges in modern IP transport networks

AI-enabled applications will boost network traffic

AI requires vast amounts of data both to train models and to perform inferencing at runtime. Moving this data around wide-area networks will strain existing infrastructure. This problem may not have presented itself yet for many service providers, but the smart ones are looking ahead.

> *Today, AI is not much in terms of traffic … but we know we are on the cusp of something that is going to be massive.*
>
> Bruno Zerbib, CTO of Orange

Omdia's *AI Network Traffic Forecast: 2022–30*, published in late 2023, predicts that upgrading existing applications with AI capabilities (e.g., image and video recognition) will lead to a significant expansion in network traffic in the coming years (see **Figure 1**). By 2030, nearly two-thirds of network traffic will involve AI, and video and image content will have grown rapidly, from 63 exabytes (EB) of global network traffic per month to 1,226 EB. These trends drive forward historic 25% year-on-year traffic increases.

As Omdia Chief Analyst Brian Washburn noted in his recent AI traffic forecast report: "Preparing networks for a rising tide of AI traffic demands more than capacity planning. Changes in network investment focus, service-level commitments, and much more are on the horizon."

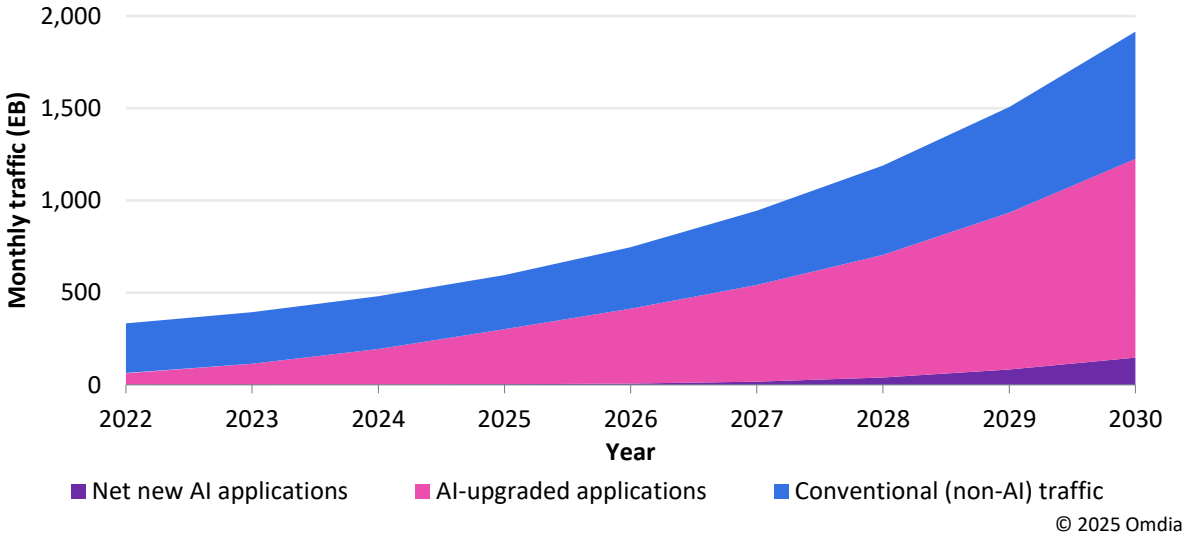
In Omdia's forecast for AI traffic through 2027, B2B applications related to video and image analysis, including responses, will be the main AI activities driving incremental traffic in the early phase.

Examples include surveillance and physical security, worker safety, quality control, health assessments, and people/object recognition. Devices will capture visual data to send upstream for an AI platform to evaluate and make decisions. The traffic will be predominantly upstream, and the downstream traffic will be limited to the machine learning (ML) model's output (an action that a downstream software system must execute).

Published February 2025

Conversely, from 2028 to 2030, growth will be led by consumer-oriented AI collaboration and entertainment traffic and by enterprise-related XR, 3D/holographic, and sensory data. By its nature, this traffic will be more downstream heavy (from the cloud to the user device).

Figure 1: Projected global network traffic growth, 2022–30



Source: Omdia AI Network Traffic Forecast: 2022–30

Challenges faced by existing IP networks

Because of the rigidity and complexity of most service providers' existing IP transport networks, they face several challenges in the AI era.

High latency

Many factors, such as long fiber distances and the location of service terminations, can induce service latencies in IP networks. Long physical distances as a result of locating service terminations and data center workloads further away from users and service endpoints can increase latencies and propagation delays.

Lack of observability

Many networks lack the telemetry and analytics to support intelligent, closed-loop automation and assure an excellent customer experience.

Lack of integrated security

IP networks will need end-to-end integrated security against all distributed-denial-of-service (DDoS) attacks to block malicious users and attacks in the AI and quantum era.

Published February 2025

Space and power constraints

A low-capacity IP platform that is power and space inefficient can consume valuable real estate resources. Aging IP and data center networks pose energy consumption stress in transporting and processing more AI workloads and can strain networks.

Layering and “silo-ization”

The service providers’ existing networks are layered and nonconverged. This nonconvergence and layering for mobile, residential, and business services in separate routing and optical networks can increase operational complexities and cost in managing networks and meeting stringent performance benchmarks in the AI-centric era.

The need to support edge computing

Some AI traffic requires processing at the edge for faster response times and data sovereignty. Enabling edge computing can also reduce transport costs, because traffic does not need to be backhauled to a centralized location for processing.

AI traffic to drive IP network upgrades

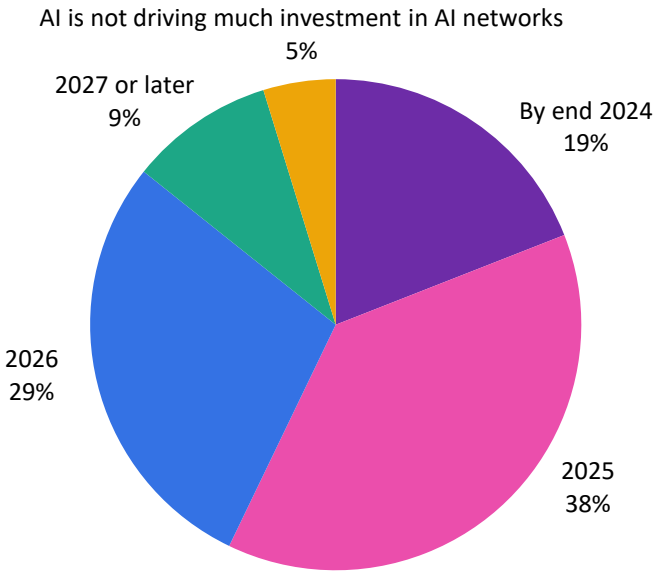
These challenges are leading many service providers to consider upgrading their IP transport networks.

> *Without a technology refresh/upgrade, service providers cannot do much on the AI front.*
>
> Haris Chughtai, Senior Technology Architect, TELUS Communications

According to a recent Omdia survey (see **Figure 2**), 20% of communications service provider (CSP) respondents expect AI to start driving additional investment in their IP transport networks by the end of 2024, 49% thought its impact would be felt in 2025, and 30% thought it would happen next year or the year after.

Published February 2025

Figure 2: When do you expect AI investments to trickle down to IP transport networks?



Source: Omdia AI for Networking and Networking for AI: Service Providers Switching and Routing Survey Results – 2024

Published February 2025

Overcoming barriers: Building smarter networks for an AI-centric era

As service providers look to prepare their IP transport networks for the AI-centric era, they must consider several factors:

- **Consolidation:** A unified and converged transport network spans access and metro and serves enterprise, home broadband, and mobile markets. By unifying what were traditionally layered, siloed, and nonconvergent silos, service providers can invest once and monetize many times over.

- **Capacity upgrades:** Service providers should consider upgrading metro networks with 400 GbE efficient routers for localized routing, network optimization, and low latency for campus, residential, and mobile xHaul services. Similarly, they should future-proof by scaling IP core networks to 800 GbE to handle massive AI workloads from data centers.

- **Flexible service distribution:** A key characteristic of AI-centric networks is the extension of services more deeply into the metro to reduce latency. Deeper service distribution in the IP metro network can be achieved by implementing network architectural and operational techniques:

  - Place edge routing closer to the user to reduce the need to send traffic to a centralized cloud or service processing node.
  - Use content delivery networks to cache copies of popular content close to users. One copy can serve several users, reducing the traffic coming from the core to the metro.
  - AI catalyzes enormous edge computing opportunities. AI traffic will hit edge nodes closer to end users rather than large, centralized cloud or remote data centers, ultimately meeting the low-latency requirements, ensuring application performance and user experience, and opening doors for new monetization opportunities.
  - Traffic engineering with Segment Routing (SRv6) is used to optimize low-latency traffic paths in metro networks and avoid congested routes.

- **Integrated end-to-end security:** Quantum computing poses a severe threat to traditional encryption techniques. Not only must IP traffic be protected with traditional protocols such as IPsec and MACsec, it will also require quantum key distribution and other postquantum security techniques. Integrated network security measures can help service providers ensure data

Published February 2025

integrity, authenticity, and confidentiality when transmitting over IP networks. They protect sensitive data against potential compromises.

How AI can benefit IP networks

In addition to generating more network traffic, AI can be leveraged in IP networks to improve performance. Examples include:

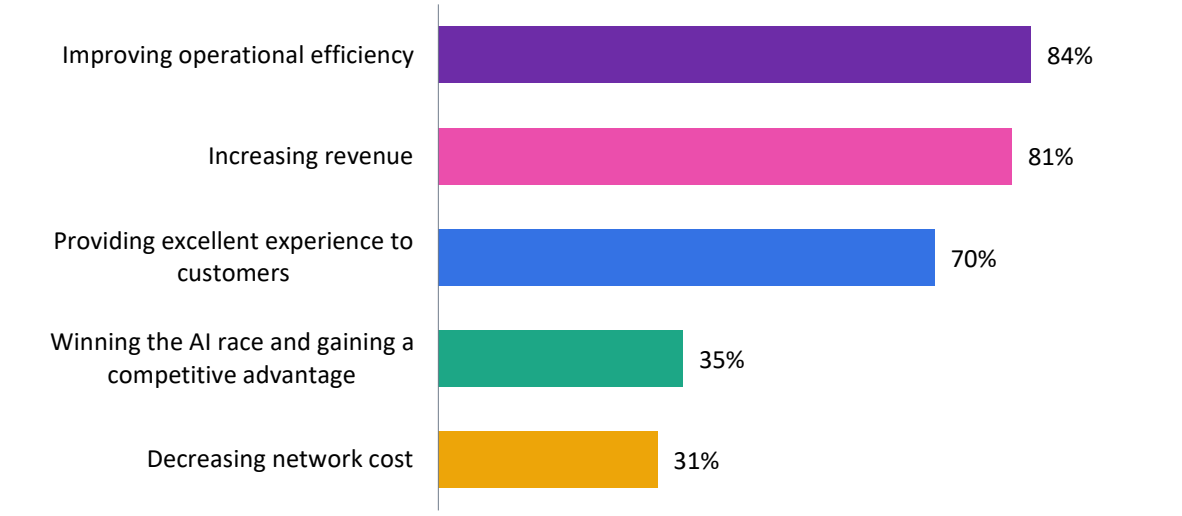
- Identifying security threats
- Predicting faults
- Enabling more intelligent routing
- Dynamically reallocating resources

According to a recent Omdia survey (see **Figure 3**), service providers expect AI-based IP operations (AIOps) to deliver three key benefits to their IP networks:

- Improve operational efficiency by increasing the level of automation and reducing the scope for manual error.
- Improve customer experience by increasing the reliability of the network.
- Increase revenue by enabling new services to be provided to hyperscalers, data center service providers, and enterprises.

Published February 2025

Figure 3: What are your top expectations for AIOps' role in the IP network?



Note: n=110

Source: Omdia, AI for Networking and Networking for AI: Service Providers Switching and Routing Survey Results – 2024

These expectations are reflected in the quote below from a Tier 1 North American service provider.

> *AI will bring operational efficiencies and better network performance, which will result in an improved user experience.*
>
> Haris Chughtai, Senior Technology Architect, TELUS Communications

Improving operational efficiency

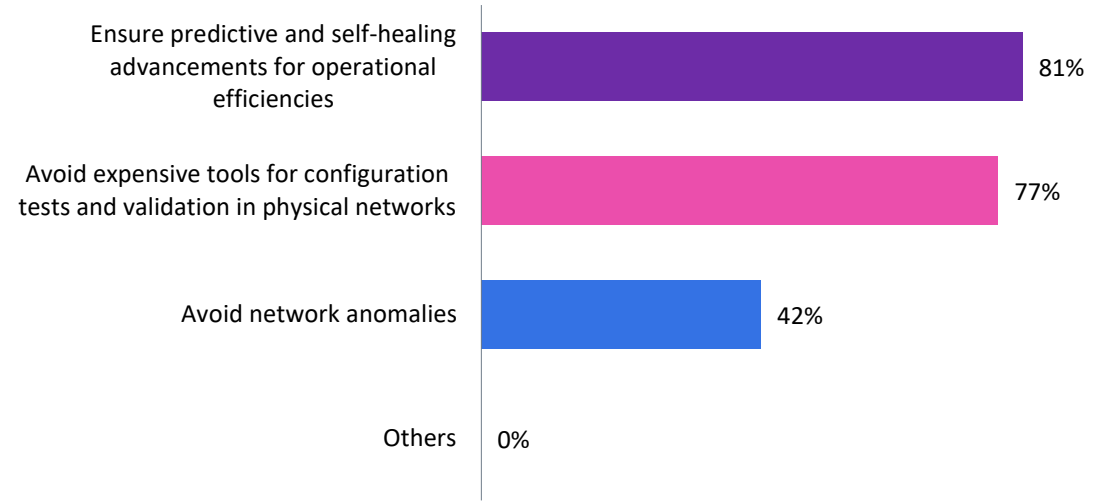
Combining real-time data and network insights into an AI-enabled management system will help service providers improve operational efficiency and simplify provisioning.

As Omdia's recent survey of service provider routing and switching professionals found (see **Figure 4**), the top drivers for AIOps are:

- Fault prediction and self-healing
- Reducing configuration testing costs

Published February 2025

Figure 4: Main drivers of AI-enabled, intent-based, self-driving autonomous networks



Note: n=110

Source: Omdia, AI for Networking and Networking for AI: Service Providers Switching and Routing Survey 2024

Increasing revenue

As well as providing connectivity between data centers and between enterprises and the cloud, service providers could potentially generate revenue from providing edge computing services themselves, as noted in an interview with Lumen Product Management.

> *AI, cloudification, and digitization are increasingly common for B2B businesses.*
>
> Jeff Ary, VP Product Management – Infrastructure Services, Lumen Product Management

The high latency (long round-trip time) of a centralized cloud can have an impact on the performance of some applications. Hosting these at the network edge instead (just as content delivery networks do today) could enhance the application performance and open new horizons for business monetization of low-latency network services.

Published February 2025

Performing AI/ML inferencing at an edge site also reduces the need to backhaul large volumes of endpoint-generated traffic (e.g., CCTV feeds). Furthermore, data security can be improved by keeping the raw data at the edge and only sending the relevant features used by the model to the centralized cloud.

Omdia analyst Diana Gorelik attended a presentation by Orange CTO Bruno Zerbib in October 2024 in which the operator emphasized a need for edge computing to optimize AI traffic handling. Zerbib's view was that with effective edge computing on telco networks, AI processing will end up in more edge computing facilities rather than on the public cloud or in data centers. However, the IP infrastructure that most service providers use today does not leverage AI in this way.

Technical tool kit for the AI era

Omdia highlights four important technologies that service providers should consider when preparing their IP networks for AI:

- **Simplified networking built with common and unified silicon architecture and system:** Selecting routers with vendor-designed unified system architecture, including next-generation ASICs (as opposed to merchant silicon) and unified network operating system, saves cost, boosts line rates, reduces latency, provides longer network life, eases upgrade, enhances security, and offers the best programmability and performance in meeting the gigantic demands of AI traffic and service requirements.

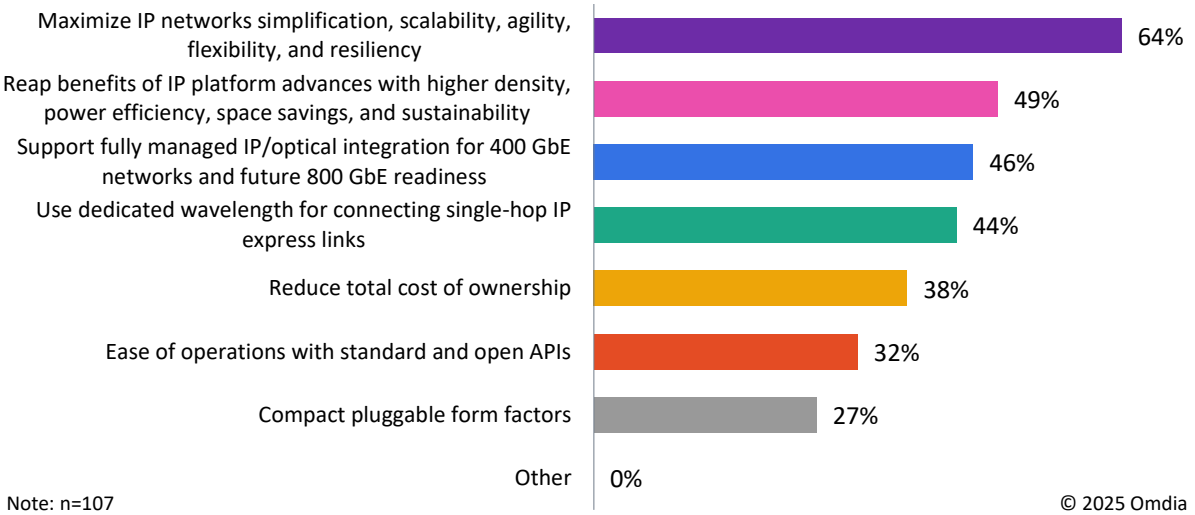
- **Converged architecture where functionality from optical transponders to line system amplifiers can be hosted as pluggables in routers:** The enduring value of colored optics in an IP router (a.k.a. IP/optical convergence) maximizes networking performance while ensuring low cost per bit. IP/optical convergence may be an inevitable choice for network modernization for the following good reasons:

  - Maximizing IP network simplification with convergence
  - Realizing IP transport network scalability, agility, and flexibility
  - Maximizing networking performance
  - Ensuring a low-cost-per-bit IP network design

- As **Figure 5** shows, when Omdia asked service providers what their top driver for implementing IP/optical-enabled digital coherent optics was, the most popular response was IP network simplification, scalability, flexibility, and resiliency.

Published February 2025

Figure 5: Top drivers for implementing IP/optical convergence



Source: Omdia Service Provider Switching and Routing Survey

- **Programmable networking with Segment Routing (SRv6) and automation to simplify traffic engineering:** SRv6 is the best choice for traffic engineering. It can cost-effectively support the robust service-level agreements (SLAs) required for mobile xHaul, enterprise, and residential applications. SRv6-based end-to-end network slicing provides the most flexible and scalable solution in the AI era. Paired with Flex Algo, it allows service providers to create differentiated SLAs, including low-latency services. It is a powerful approach that helps to manage network resources and prioritize VIP users or mission-critical services traffic, avoiding congested links. Real-time AI inferencing in fulfilling agreed-upon SLAs will enable service providers to monetize their network resources with guaranteed SLAs for various service offerings, such as VIP users, integrated security, and mission-critical assurance.

- **Observability that provides real-time and end-to-end visibility into network and service metrics:** AI-enabled observability revolutionizes IP networking and reduces operational headaches. The nexus of AI and ML in IP networking ensures intelligent, resilient, secure, and optimized IP networking performance where AI-based zero-touch IP network provisioning, WAN automation, AIOps in cloud and network integration for digital transformation, and ML-based advanced and predictive analysis will remain the center points.

Published February 2025

Conclusions: Shaping the future of IP networks

The emergence of generative and agentic AI will drive significant traffic growth over the next five years for both enterprise and consumer applications. Adding capacity to current IP transport networks will not be the most cost-effective or performant solution. Instead, network service providers must take a fresh look at their network architecture and technologies to support agile services networking.

Existing networks face limitations that make them struggle in supporting the emerging services requirements for the AI-centric era:

- The centralization of service nodes and cloud applications in distant locations for emerging AI use cases leads to high latency.
- Weak observability means operations teams cannot ensure the experience that customers demand.
- Networks are not prepared for quantum decryption.
- Equipment is bulky and power hungry.
- Access, metro, and core networks operate in silos.
- The network is not optimized to support emerging edge compute and real-time AI inferencing of AI use cases and applications.

Beyond the apparent upgrades of throughput (400 GbE for metro networks, 800 GbE for core) Omdia recommends the following key technologies to prepare IP networks for the AI era:

- Converging layers and simplifying architecture in the form of IP/optical network convergence ensures network simplification, scalability, and flexibility with low cost and high performance.
- Segment routing for traffic engineering ensures the robust SLAs required for mobile xHaul and business and residential applications in the AI era.
- Flexible placement of edge services and AI inferencing improve responsiveness.
- Network slicing enables differentiated service offerings for VIP users.

Published February 2025

- Next-generation efficient and rich-feature-set ASICs enhance security, programmability, scalability, line-rate performance, and form factors for routers.
- IP/optical network convergence ensures network simplification, scalability, flexibility, and resiliency.
- AIOps offers operational simplicity to manage and automate non-AI and future net-new AI traffic workloads.

As the quote below highlights, to win in the AI era, the next-generation IP architecture must be software defined, programmable, open, dynamically scalable, highly efficient, and highly performant.

> *The MPLS transport network is evolving into a next-generation architecture, embracing programmability, open standards, and software-defined capabilities to meet modern connectivity demands. This transformation enables dynamic scalability, efficient resource utilization, and seamless integration with networks, supporting the high-performance, low-latency, and agile requirements of the emerging digital ecosystem.*
>
> Head of Data & Network Planning, Tier 1 Middle Eastern service provider

Published February 2025

Appendix

Methodology

This paper is based on the author's knowledge of the telecom industry and makes use of numerous Omdia publications that are part of the Service Provider Routing and Switching Intelligence Service.

Further reading

AI Network Traffic Forecast: 2022–30 (November 2023)

AI for Networking and Networking for AI: Service Providers Switching and Routing Survey Results – 2024 (October 2024)

Service Provider Routing and Switching Technology Survey Results – 2023 (August 2023)

Author

Sameer Malik
Senior Principal Analyst, Service Provider Routing & Switching
askananalyst@omdia.com

About Cisco

Cisco™ Agile Services Networking is an architecture to power AI connectivity and customer experiences that enables service providers to monetize the delivery of assured services and networking.

With Cisco, organizations can monetize services deployed closer to end-user demand using a network architecture optimized for:

- Intelligent service delivery
- Simplified networks that converge network layers and services
- Assured experiences with resilient networks and services enabled by AI-powered automation, observability, and security

Get in touch

www.omdia.com
askananalyst@omdia.com

Omdia consulting

Omdia is a market-leading data, research, and consulting business focused on helping digital service providers, technology companies, and enterprise decision makers thrive in the connected digital economy. Through our global base of analysts, we offer expert analysis and strategic insight across the IT, telecoms, and media industries.

We create business advantage for our customers by providing actionable insight to support business planning, product development, and go-to-market initiatives.

Our unique combination of authoritative data, market analysis, and vertical industry expertise is designed to empower decision-making, helping our clients profit from new technologies and capitalize on evolving business models.

Omdia is part of Informa TechTarget, a B2B information services business serving the technology, media, and telecoms sector. The Informa group is listed on the London Stock Exchange.

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 your company identify future trends and opportunities.

Copyright notice and disclaimer