

Meeting the Challenges of Operating Infrastructure in the AI Era

Author: Roy Illsley

October 2026

In partnership with:



This Omdia White Paper was commissioned by Park Place Technologies.

Contents

Summary	3
Catalyst	3
Omdia view	3
Evolving architecture to support AI growth	5
The rise of AI factories	5
AI factory challenges for mid-market CIOs	7
GPUs introduce new and distinct challenges	8
AI factories require a change in operational management	9
Maintaining all the skills needed to operate an AI factory requires a different approach ..	9
Real-time management requires IT and OT convergence at the data layer	10
Asset visibility is key in an AI factory	12
A leading higher education institution specializing in AI	13
The environment	13
The challenge	13
The solution	14
The results	14
Key takeaway	15
Managing AI factories requires new skills, new thinking, and strategic partnerships	15
Appendix	16
Authors	16

This Omdia White Paper was commissioned by Park Place Technologies.



Summary

Catalyst

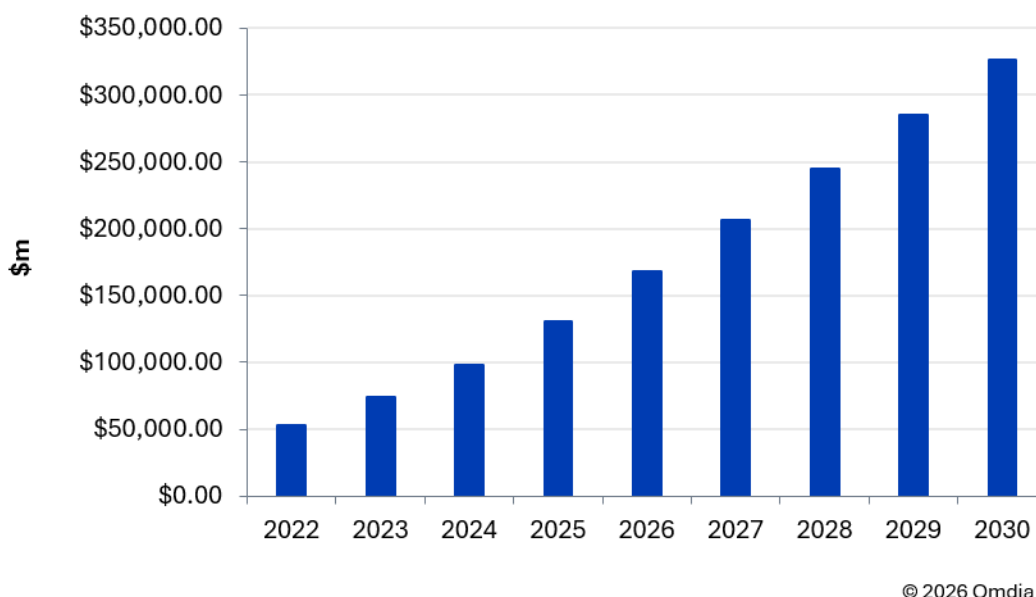
As technology evolves at an ever faster pace, the ability of chief information officers (CIOs) to develop, maintain, and operate technology environments becomes an even greater challenge. This challenge is amplified by the fact that, according to Omdia's *IT Enterprise Insights: IT Spending & Sourcing – 2026* survey, only 31% of respondents reported that their IT budget increased significantly, meaning 69% reported that their IT budget increased only slightly, remained flat, or decreased in 2026 compared with 2025. As a result, CIOs must plan and develop the next technology operating model that can deliver new technology while also maintaining existing services for the artificial intelligence (AI) era.

Omdia view

The rise of generative and agentic AI shows that by 2030, according to Omdia, the AI software market will be worth \$369bn (see **Figure 1**), an increase from \$131bn in 2025. In fact, 44% of organizations reported that they have adopted AI, according to *IT Enterprise Insights: AI – 2026*. This growth in investment will increase the scrutiny of how this money is invested, and how this new architecture will be managed by already overstretched IT departments.

This Omdia White Paper was commissioned by Park Place Technologies.

Figure 1: Omdia's forecast for AI software market



Source: Omdia

To deliver the AI needed by businesses, the global data center landscape is undergoing a period of profound structural change, driven by escalating computing demand, a critical need for operational resilience, and the rapid rise of high-density AI workloads. Omdia's data shows that data center capex spending is on track to reach \$1.6tn by 2030, and growing at a 17% compound annual growth rate (CAGR) from 2025 till 2030. Omdia's forecast considers the actual order pipeline and demand, both of which were strong in 2025. Enterprise customers are investing in new AI factories that are based on new architectures, which include new technologies (e.g., GPUs and liquid cooling) that require new operating processes and skills many enterprises currently lack. When this is combined with the current global supply constraints on AI chips, these AI factories frequently must mix legacy clusters with cutting-edge accelerators, leading to highly complex hybrid infrastructure.

One of the key drivers of enterprise customers investing in data centers is the rise of digital sovereignty, particularly in terms of sovereign AI. In fact, 2026–27 will be pivotal for AI development—and hence sovereign AI—as regulations, government investment, and geopolitics converge to make sovereignty a critical risk management strategy.

This Omdia White Paper was commissioned by Park Place Technologies.

Many governments are betting big on sovereign AI: for example, the UK's £500m (\$673m) investment in a new Sovereign AI Unit to boost AI capabilities; Saudi Arabia's HUMAIN sovereign AI venture backed by around \$100bn in mostly government funding; and Singapore channeling over S\$1bn (\$788m) into AI development, with S\$500m dedicated to securing high-performance compute.

Governments are demonstrating a willingness to implement sovereign AI strategies that they hope will drive economic growth and maintain global competitiveness while safeguarding their sensitive data. Omdia expects the term "sovereign generated data" to become more common as customers realize that using AI to produce new data sets from existing corporate data sets is becoming an AI use case and must be considered from the perspective of who owns the IP and how to protect it.

Most countries have some legal frameworks that requires certain data to be stored in the country. For example, in April 2023, the Vietnamese government passed its first comprehensive data protection law, Decree 13/2023/ND-CP on the protection of personal data. This shows that sovereignty is more than just a European discussion, and globally, enterprise customers are considering where to invest in IT capability to meet these sovereignty requirements.

However, Omdia believes the future will be a hybrid model with cloud and on-premises infrastructure operating seamlessly, but with strict segmentation and compliance being applied. To operate in this complex hybrid environment, organizations will need to partner with experts and ensure that this new architecture and traditional architecture do not become two separate operational processes.

Evolving architecture to support AI growth

The rise of AI factories

Organizations face three choices for AI deployment: a cloud-based approach, a fully on-premises approach for sovereignty and privacy reasons, or a hybrid model. The hybrid approach appears to be gaining the most traction, with AI-related workloads witnessing the largest shift from cloud to on-premises—65% of organizations reported adopting sovereign environments for AI workloads in 2026–27, according to Omdia's *IT Enterprise Insights: IoT, Cloud, and Sustainability – 2026* survey. Notably, this trend runs counter to most other workloads, which continue migrating from on-premises to the cloud. For example, 73% of customer relationship management (CRM) workloads are expected to run in the cloud.

This Omdia White Paper was commissioned by Park Place Technologies.

Overall, the current distribution shows 36% of workloads run on-premises, 36% in the cloud, and 28% run as software-as-a-service (SaaS).

This suggests a growing market for specialized AI-optimized data center facilities equipped with high-density GPU clusters, advanced cooling systems capable of handling the thermal output of AI processors, and network architecture designed for the massive data movement requirements of AI training and inference operations.

The development of on-premises AI factories represents a fundamental architectural shift. These AI factories act as the hybrid bridge between on-premises and cloud infrastructure for many organizations. Omdia believes the growth of AI factories will transform data center operations and management, introducing new processes and practices.

Figure 2 shows the top three risks that respondents in Omdia’s *AI Market Maturity Survey – 2026 Data Tool* identified as important considerations when adopting AI. Security and privacy rank first and third, respectively, reinforcing the trend of AI workloads migrating from public cloud to on-premises environments.

Figure 2: Top three risks ranked as most important considerations for AI adoption



© 2026 Omdia

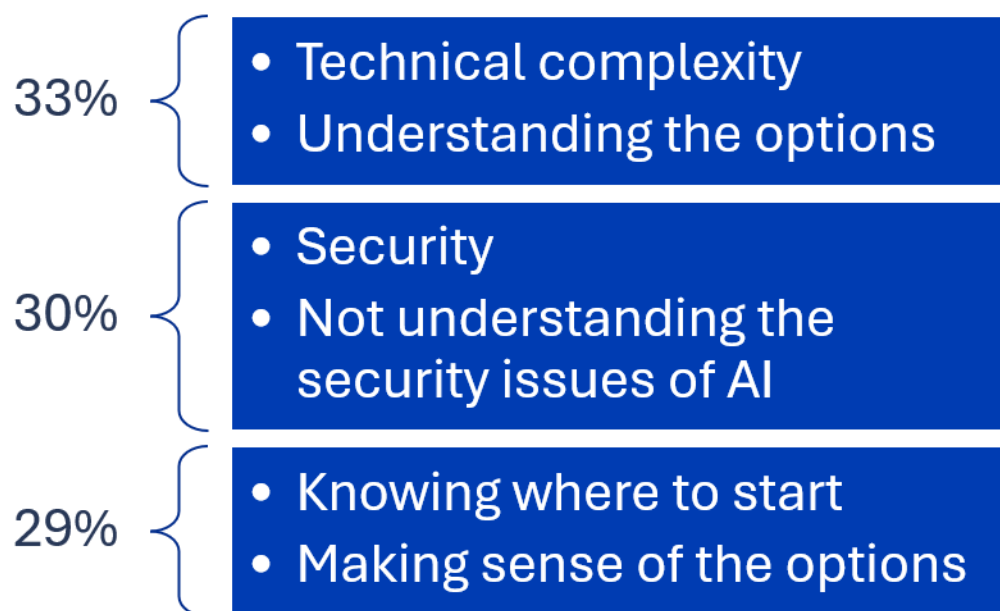
Source: Omdia

This Omdia White Paper was commissioned by Park Place Technologies.

AI factory challenges for mid-market CIOs

Mid-market customers are increasingly turning to private cloud solutions to deploy AI capabilities due to rising concerns over data security, compliance, and unpredictable cloud costs for AI workloads. However, these organizations face specific challenges when adopting AI factories.

Figure 3: Top three challenges mid-market customers face when adopting AI



Notes: n=383

© 2026 Omdia

Source: Omdia

The top concern is the technical complexity of the AI factory, specifically, understanding the available options and differentiating between vendors' solutions. While this may appear to be a purchasing dilemma, it quickly becomes an operational challenge: How do the infrastructure and software layers work together? What actions are required to maintain system operations?

Interestingly, scaling is not a major concern, with only 17% of customers reporting the ability to scale AI as a challenge. Omdia considers this evidence that customers view on-premises solutions favorably because they provide better cost control, preventing the budget overruns that can occur in cloud environments.

This Omdia White Paper was commissioned by Park Place Technologies.

The second concern is security and understanding how to ensure that the environment and all its components are not compromised or used in a way that exposes the business to increased risk. This concern further demonstrates a knowledge gap about how AI systems are built and operate, reinforcing Omdia's view that customers need guidance from expert partners.

This knowledge gap becomes even more apparent with the third concern: knowing where to start. These findings reveal that customers' limited understanding about AI systems and AI factories is the primary barrier, a stark contrast to their clear grasp of AI's benefits and value.

For example, mid-market customers identify workforce productivity as the prime motivation for AI investment, with 40% citing it as their top use case. Additionally, Omdia's *AI Market Maturity Survey – 2026 Data Tool* found that 38% of all respondents—the largest group—expect to spend 11% to 25% more in 2026 on AI in 2026 compared with 2025.

GPUs introduce new and distinct challenges

To successfully deploy GPUs in an AI factory, IT departments must transition from isolated server management to an integrated, cluster-wide infrastructure management.

- **Orchestration and containerization:** Deploying GPUs individually is inefficient. IT teams rely on platforms like Kubernetes, paired with specialized plug-ins (e.g., the NVIDIA GPU Operator) to automate container deployment, scaling, and health monitoring.
- **High-performance networking:** GPUs in an AI factory cannot operate in isolation; they must constantly exchange data during distributed training operations. Deploying ultra-low latency fabrics, such as InfiniBand or RDMA over Converged Ethernet (RoCE), is essential to prevent networking bottlenecks.
- **Optimized storage architecture:** GPUs process data faster than traditional storage can deliver it. Enterprise IT must deploy high-throughput, parallel file systems (e.g., NVMe-based data lakes) to ensure the compute engines are never starved of data.

Introducing these specialized accelerators into a standard corporate data center creates friction across several vectors.

- **Power and cooling:** Modern GPU servers, such as those hosting NVIDIA H100s or Blackwell chips, draw unprecedented amounts of power. Standard enterprise racks are built for 5–10kW, whereas AI factory racks require 40–100 kW or more, forcing IT departments to adopt complex liquid cooling systems.

This Omdia White Paper was commissioned by Park Place Technologies.

- **Specialized skills gap:** Managing GPU clusters requires specialized knowledge of CUDA libraries, AI frameworks (e.g., PyTorch and TensorFlow), and high-performance clustering fabric. Standard IT system administrators face a steep learning curve.
- **Cost management and utilization:** GPUs represent massive capex, meaning IT departments must closely monitor idle time to prevent millions of dollars from being wasted on underutilized compute resources.
- **Security and compliance:** Funneling massive, proprietary corporate datasets into a centralized GPU factory significantly increases the attack surface. IT must ensure strict access controls, data encryption, and regulatory compliance measures, all without creating bottlenecks that slow down data scientists.
- **Operational transformation:** Ultimately, supporting GPUs in an AI factory fundamentally shifts the enterprise IT department from a traditional service provider into a high-performance computing (HPC) operator, requiring massive upgrades to physical infrastructure, software stacks, and internal expertise.

AI factories require a change in operational management

Maintaining all the skills needed to operate an AI factory requires a different approach

The difference between delivering traditional IT services and AI services extends far beyond the infrastructure and software being used—the expectations of the customers also increase. For IT teams, this means operational processes must change. Data centers serving such customers must provide consistent management interfaces, unified observability, and reliable connectivity across distributed environments. This drives demand for sophisticated orchestration platforms, multi-cloud networking solutions, and management tools that abstract underlying infrastructure complexity.

For example, the new AI servers that are being developed are increasingly deployed at rack scale, introducing new management challenges. Existing OT software typically does not operate at rack scale, so new racks used for these power-dense AI servers include rack management capabilities that cover the different aspects, such as leak detection and auto-valve shutdown, and three-phase power balancing at the rack.

This Omdia White Paper was commissioned by Park Place Technologies.

This highlights another area where customers face critical challenges, namely, in the skills and recruitment of the appropriate technical expertise. With the infrastructure continually evolving—liquid cooling, rear door heat exchangers, and battery energy supply systems—these skills require constant updating to maintain currency. This becomes critical to ensure that the service delivery resiliency that is being built into the infrastructure can be operated correctly.

IT department processes and roles have evolved over time to ensure traditional data centers provide appropriate operational management. However, the rise of AI, and of agentic AI, will eventually transform the operational structure, roles, and responsibilities of IT departments.

Organizations must plan for this transformation now, defining clear roles for technology in the business as AI enables many more end-users to become operationally self-sufficient. This means that the current approach to technology operational management, based on organizational structures that evolved by segmenting activities into skill-based teams and departments reporting to C-suite executives, must change.

Historically, activities were segmented by skills and knowledge. In an agentic AI era, these will no longer be the defining factor for activity allocation. As a result, organizations must plan how to retain the skills the business needs, exploit AI capabilities effectively, and maintain critical infrastructure operational management expertise.

Ultimately, processes will be redefined and optimized based on how they best deliver business benefits, fundamentally changing traditional roles and responsibilities. This will create a challenge for CIOs because how this change will manifest remains unclear. Omdia believes it could create a separation:

- Generalists will be the predominant users and operators of new agentic AI tools
- Deep technical specialists will be required for specific activities

The question CIOs must answer is: Where will the skills needed to perform these new activities come from—internal personnel or external partners?

Real-time management requires IT and OT convergence at the data layer

One critical consequence of the change mentioned earlier is that these new AI factories will require real-time management capabilities. However, many data center operators are

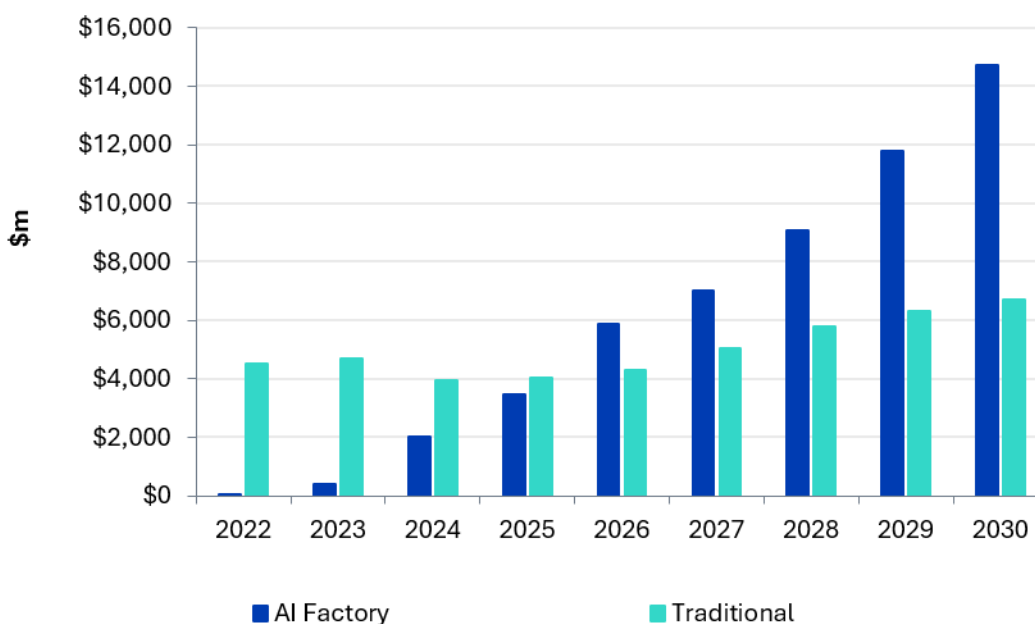
This Omdia White Paper was commissioned by Park Place Technologies.

encountering difficulties with using the current management and monitoring capabilities in real-time, because the activities are typically segmented into different domains—IT and OT.

This challenge is compounded by the vendor landscape, which evolved to serve these domains separately. The IT equipment manufacturers’ solutions, known as “white space” solutions, are used for managing the racks where the compute, storage, and network assets are housed. Meanwhile, the OT equipment manages the power and cooling of the physical data center.

Figure 4 shows how the OT vendors’ software has changed to meet the challenges of the new AI factories, and in 2026, Omdia expects AI factories to become the largest market for these OT software solutions.

Figure 4: How OT software is shifting to manage AI factories



© 2026 Omdia

Source: Omdia

The OT management and monitoring solutions cover incoming power and distribution to racks and other consumers, as well as thermal management of the data center—collectively known as “gray space.” The emergence of power-dense racks is creating a new discipline: rack management, which represents a hybrid capability bridging IT and OT domains.

This Omdia White Paper was commissioned by Park Place Technologies.

The OT software market is highly fragmented, in terms of the software used. For example, building management systems (BMS) typically control office power and site-level cooling management. On the other hand, data center infrastructure management (DCIM) solutions control site power distribution, non-office power consumption, and mechanical aspects of cooling systems. This results in a confusing array of products with overlapping capabilities. Consequently, most data center operators deploy multiple tools across different teams to manage the entire facility, creating silos, inefficiencies, and gaps in real-time visibility.

Meanwhile, IT “white space” solutions, such as artificial intelligence for IT operations (AIOps) and cloud operations (CloudOps), manage IT applications and workloads.

This fragmented approach is not only inefficient, but it also exposes gaps in the domain-specific tools, creating organizational knowledge gaps. For example, operators cannot easily attribute power and cooling costs directly to specific workloads running on servers in individual racks.

While transforming power and cooling management for data centers remains a long-term ambition, Omdia believes data center operator demand will drive market change. This change will likely emerge first in AI financial management, where power and cooling costs represent a significantly greater proportion of total workload costs.

IT management and monitoring capabilities have evolved faster than OT systems, and AI-driven tools are now common. However, the critical challenge is combining data from both domains to predict failures before they occur.

To achieve this predictive capability, IT and OT domains must share data, adopt standard definitions, and support process integration regardless of the specific management and monitoring software deployed.

Asset visibility is key in an AI factory

The complexity and change CIOs must adopt is forcing organizations to transition from human-centric, human-constrained operations to automated, predictive systems capable of preventing catastrophic compute failures. This shift becomes increasingly critical as AI factories emerge as the dominant form of enterprise computing. This is because, in AI factories, organizations must transition from traditional hardware break-fix workflows to a proactive ecosystem powered by automated machine learning algorithms. This requires:

- Establishing environment-specific baselines to minimize false alarms
- Detecting performance degradation before component failure

This Omdia White Paper was commissioned by Park Place Technologies.

- Integrating monitoring across multiple vendors and architectures.
- Covering storage platforms, standard servers, networking equipment, and high-density compute environments.
- Monitoring all AI factory components, including liquid cooling, GPUs, and so on.

The stakes are extraordinarily high. If a single node in a 10,000-GPU cluster fails due to a memory or hardware fault, the entire parallel training job can crash, corrupting active workloads and costing hundreds of thousands of dollars in wasted power and delayed time-to-market.

This raises a critical question: How can existing IT and OT operational processes, teams, and software deliver this complex, holistic management and monitoring efficiently and effectively?

The challenge extends beyond traditional infrastructure. AI factories depend on hyper-fast, low-latency networking fabrics (e.g., InfiniBand or RoCE) using massive arrays of transceivers and optical cables. These require mature network management capabilities fully integrated into IT and OT processes.

A leading higher education institution specializing in AI

The environment

The speed of change in AI research requires organizations to evolve their IT environment to ensure they remain able to develop cutting-edge AI capabilities. As a result, the customer's IT environment became complex and consisted of different vendor technologies that included both CPUs and GPUs. The customer's environment consisted of Supermicro servers comprising 55 GPUs (including NVIDIA A100s), 34 CPU servers, 18 cluster nodes, and Cisco Catalyst 9300 switches.

The challenge

The university lacked a coherent, consistent approach to technology management. It relied on an independent software vendor for portions of its technology stack in another part of the university. However, this vendor had incomplete visibility across the environment,

This Omdia White Paper was commissioned by Park Place Technologies.

could not investigate, resolve, and report incidents beyond its limited scope. This meant that problems were not diagnosed correctly, resulting in sub-optimal resolutions.

The solution

The university required a non-OEM partner capable of supporting a complex, mixed IT environment, including GPUs, and providing the ability to identify, remediate, and resolve incidents for the university.

Park Place Technologies met these requirements as one of the few non-OEMs offering GPU support and comprehensive automated monitoring. Its ParkView Automated Support™ platform delivers multi-platform, multi-vendor monitoring and management.

The results

During initial engagement, Park Place Technologies immediately demonstrated its value. When the university experienced a recurring problem, Park Place Technologies investigated device-level health across the entire system and identified a cabling issue. By using a solution and provider that could deliver device-level visibility, automated ticket routing and resolution, and comprehensive hardware support, the university gained greater operational confidence in its environment.

This Omdia White Paper was commissioned by Park Place Technologies.



Key takeaway

Managing AI factories requires new skills, new thinking, and strategic partnerships

The rise of new infrastructure, operating procedures, and the mix of equipment in AI factories means organizations must consider how they are going to manage these systems. When the new AI factory is being used, organizations will be adopting a hybrid approach to infrastructure, meaning it is not just on-premises but also utilizes cloud computing.

When this is combined with the constantly changing technologies and the need to operate and integrate them into the operating environment at speed, current approaches will not deliver the required outcomes.

Current approaches to managing and operating technology must change. Organizations must carefully consider the roles and responsibilities of their employees, factor in the use of external third parties, and determine how they fit into the operating processes.

This Omdia White Paper was commissioned by Park Place Technologies.

Appendix

Authors

Roy Illsley, Chief Analyst, Enterprise

Roy.illsley@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.

Get in touch

www.omdia.com
askananalyst@omdia.com



Copyright notice and 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. 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.