

AI Infrastructure Market to Hit USD 997.9 Billion by 2035, Registering 19.9% CAGR

AI infrastructure Market was valued at \$135.8 billion in 2024 and is projected to reach \$997.9 billion by 2035, growing at a CAGR of 19.9% from 2025 to 2035.

INDORE, MADHYA PRADESH, INDIA, April 19, 2025 /EINPresswire.com/ --

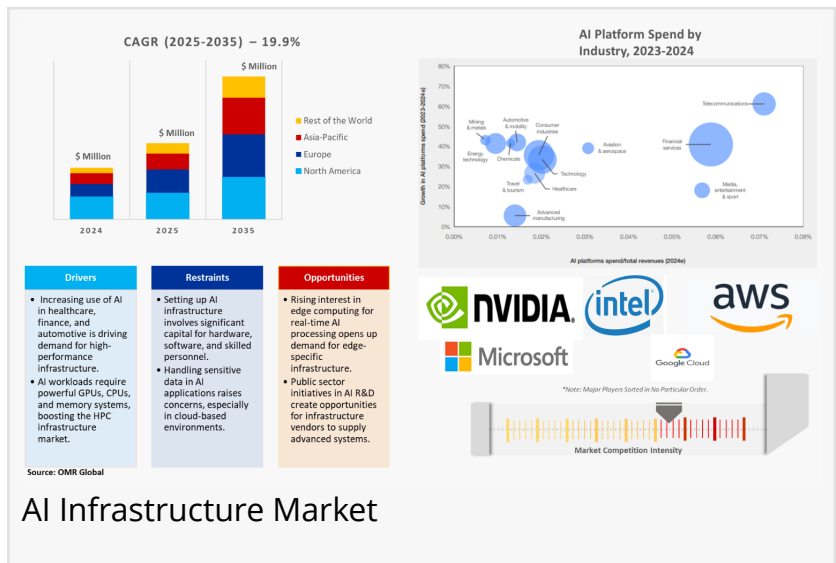
The [AI infrastructure market](#) growth is driven by the increasing AI integration in different sectors, compute infrastructure, surge in Gen AI, rise in cloud-native AI deployment, and others. AI is advancing at a rapid pace, fueled by significant developments in

computing power, data availability, and sophisticated algorithms in areas such as natural language processing (NLP), computer vision, and generative AI. Moreover, AI investments vary across industries, according to their needs and digital maturity. Technology firms such as NVIDIA, Google Cloud, Microsoft, and others are leading in AI infrastructure development, heavily investing in AI chips, data centers, and R&D to create applications for AI adoption across other sectors. For instance, in March 2025, Nvidia plans to invest in US-made chips and electronics over the next four years. The company expects to spend around half a trillion dollars on electronics.



Get 15% off on Global Market Reports until May 31st!"

Orion Market Research



Click to get a sample pdf (Including Full TOC, Graphs & Charts, Table & Figures) @ <https://www.omrglobal.com/request-sample/ai-infrastructure-market>

Market Dynamics

Rise in Cloud-native AI deployment

There is a shift toward cloud-native AI where AI models, applications, and pipelines are developed, deployed, and scaled within cloud environments. It has increased the need for

flexible, scalable, and high-performance AI infrastructure. Cloud Native technologies empower organizations to build and run scalable applications in modern, dynamic environments such as public, private, and hybrid clouds. Containers, service meshes, microservices, immutable infrastructure, and declarative APIs exemplify this approach. Cloud Native Artificial Intelligence (CNAI) allows the construction of practical systems to deploy, run, and scale AI workloads. CNAI solutions address challenges AI application scientists, developers, and deployers face in developing, deploying, running, scaling, and monitoring AI workloads on cloud infrastructure. Leveraging the underlying cloud infrastructure's computing (e.g., CPUs and GPUs), network, and storage capabilities, as well as providing isolation and controlled sharing mechanisms, accelerates AI application performance and reduces costs.

For instance, Kubernetes is an orchestration platform that can be used to deploy and manage containers, which are lightweight, portable, self-contained software units. AI models can be packaged into containers and then deployed to K8s clusters. Containerization is especially crucial for AI models because different models typically require different and often conflicting dependencies. Isolating these dependencies within containers allows for far greater flexibility in model deployments. CN tooling allows for the efficient and scalable deployment of AI models, with ongoing efforts to tailor these for AI workloads specifically.

Government Initiatives Fueling AI Infrastructure Demand

Government AI strategies have a crucial role in the growth of this market by driving large-scale investment, deployment, and regulation of AI technologies. National AI programs and regulatory frameworks across countries such as the US, China, India, and EU member states are prioritizing the development of AI ecosystems, which includes building robust cloud and on-premises infrastructure, data centers, and high-performance computing (HPC) capabilities. These strategies are fueling demand for AI chips, servers, storage solutions, and networking systems to support government-led AI adoption in public services, defense, smart cities, healthcare, and more. For instance, the EU AI Act assigns applications of AI to three risk categories. First, applications and systems that create an unacceptable risk, such as government-run social scoring of the type used in China, are banned. Second, high-risk applications, such as a CV-scanning tool that ranks job applicants, are subject to specific legal requirements. Lastly, applications not explicitly banned or listed as high-risk are largely left unregulated. This leads to increased deployment of sovereign clouds and hybrid infrastructure models.

Market Segmentation and Growth Areas

Compute Segment to Lead the Market with the Largest Share

The computing market which includes CPUs, GPUs, and FPGAs is increasing as it is the core enabler of AI workloads. Organizations are undertaking increasingly sophisticated AI projects such as orchestrating multiple LLMs (and non-LLMs), across diverse frameworks, to deliver compound AI and agentic applications. AI workloads span data processing, training, tuning, inference, and serving with models of widely varying sizes and end application user requirements. Delivering these AI applications in production, customers need to be as efficient

as possible in their GPU utilization and cloud expenditures while meeting these requirements. For instance, with Google Cloud's Anyscale, customers can optimize and run models and workloads on the efficient hardware for the task. A single cluster can be made up of hundreds of machines, some with TPUs, and some GPUs, to meet specific application requirements while lowering costs. Additionally, Anyscale supports leveraging Spot, On-Demand, or fixed Capacity Reservations as it runs AI workloads - optimizing for price, availability, and efficiency.

Order Your Report Now For A Swift Delivery: https://www.omrglobal.com/buy-now/ai-infrastructure-market?license_type=license-single-user

Cloud Service Providers: A Key Segment in Market Growth

The Cloud Service Providers segment is experiencing the fastest and most significant growth due to huge capital expenditure. In January 2023, Microsoft entered the third phase of its strategic partnership with OpenAI through a multi-year, multibillion-dollar investment intended to accelerate AI advancements and expand global access to AI technologies. This extended collaboration focuses on scaling AI supercomputing systems, advancing AI research, and deploying OpenAI's models through Microsoft's consumer and enterprise products. Microsoft's Azure platform, serving as the exclusive cloud provider for OpenAI, will support all OpenAI workloads from research to product deployment and API services reinforcing Azure's position as a global leader in AI infrastructure. This move ensures that organizations and developers globally can build and run powerful AI applications using models with enterprise-grade security and scalability. Since the partnership began in 2016, Microsoft has continuously strengthened Azure's capabilities as an AI supercomputing platform.

Challenges: High investment costs for AI infrastructure

One of the restraints in this market is high investment costs. Investment in AI infrastructure is expensive. The costs associated with trying to develop AI applications and capabilities on traditional IT infrastructure can be even more costly. Generative AI is already driving up the cost of computing. The cost of developing AI can be challenging for any company, especially small and mid-sized ones. The cost of AI software development is influenced by many factors such as types of AI solutions, project complexity, AI development tools and infrastructure, type and amount of data needed, algorithm training needs, regulatory and ethical compliance, deployment and maintenance, and others.

Opportunity: Public Sector Push Accelerates Demand for AI Infrastructure

Public sector initiatives, such as the US National Science Foundation's National Artificial Intelligence Research Resource (NAIRR) pilot, are creating significant opportunities for AI infrastructure vendors. The NAIRR pilot, launched in collaboration with 10 federal agencies and 25 private, nonprofit, and philanthropic organizations, aims to democratize access to critical AI resources including advanced computing, datasets, models, and training for US-based

researchers and educators. This initiative lays the foundation for nationwide AI research and development. Vendors supplying GPUs, cloud platforms, software stacks, and support services stand to benefit as government-backed programs continue to prioritize AI-driven growth. These cross-sector collaborations fuel scientific advancement and also open new markets for infrastructure providers, positioning them to play a key role in shaping the future of AI in the US.

Regional Outlook

The global AI infrastructure market is further divided by geography, including North America (the US and Canada), Asia-Pacific (India, China, Japan, South Korea, Australia and New Zealand, ASEAN Countries, and the Rest of Asia-Pacific), Europe (the UK, Germany, France, Italy, Spain, Russia, and the Rest of Europe), and the Rest of the World (the Middle East & Africa, and Latin America).

Empowering Growth Through AI: India's Blueprint for a Digital Future

India is emerging as a global hub for AI driven by strong government backing and a vision to democratize technology under the AI for All initiative. In 2024, the Government of India launched the IndiaAI Mission which highlighted this ambition, focusing on seven foundational pillars including compute infrastructure, startup funding, skill development, and the creation of safe, trusted AI solutions. This policy push has unlocked market opportunities across key sectors. Through initiatives such as the Digital India Bhashini platform and the IndiaAI Innovation Centre, AI applications tailored to Indian languages and use cases in healthcare, education, and finance are being actively developed. Over 67 proposals for foundational and sector-specific models have already been submitted, with support in funding and infrastructure, including GPU access. In parallel, the government is heavily investing in skill-building through programs such as FutureSkills PRIME, IndiaAI Fellowships, and collaboration with academia and tech giants such as Meta. These efforts are laying the groundwork for a scalable, inclusive AI ecosystem that transforms domestic markets and also enhances India's role in the global AI landscape.

North America Region Dominates the Market with Major Share

The AI infrastructure market in North America is experiencing rapid growth owing to increasing government support and rising demand for advanced AI technologies, driven by innovations in cloud computing, data centers, and high-performance computing (HPC). According to International Energy Agency (IEA) investment in new data centers has surged over the past two years, driven by growing digitalization and the uptake of AI, which is expected to continue accelerating. Much of the spending is concentrated in the US, where annual investment in data center construction has doubled in the past two years alone, although other major economies, such as China and the European Union, are also witnessing an increase in activity. In 2023, overall capital investment by Google, Microsoft, and Amazon, which are industry leaders in AI adoption and data center installation, was higher than that of the entire US oil and gas industry

totaling around 0.5% of US GDP.

Market Players Outlook

The major companies operating in the global AI infrastructure market include Amazon Web Services, Inc., Google Cloud, Intel Corp., and NVIDIA Corporation, among others. Market players are leveraging partnerships, collaborations, mergers, and acquisition strategies for business expansion and innovative product development to maintain their market positioning.

Recent Developments

- In September 2024, Microsoft announced a \$1.3 billion investment in Cloud and AI infrastructure supporting inclusive growth through technology and skilling programs in Mexico. The plan included the AI National Skills initiative, aimed at training 5 million people, expanding connectivity, and helping SMBs adopt AI. Microsoft also supports sustainability efforts through carbon removal and water replenishment projects and expands its AI for Good initiatives in health and connectivity. These moves reinforce Microsoft's commitment to inclusive growth, innovation, and responsible AI development in Mexico.
- In September 2024, BlackRock, GIP, Microsoft, and MGX announced a \$100 billion investment partnership to boost American AI competitiveness and meet the increasing demand for energy infrastructure. It will invest in new data centers and energy infrastructure, primarily in the US, supporting AI innovation and economic growth. NVIDIA supports the partnership, offering expertise in AI data centers and factories. The partnership includes \$30 billion in private equity capital and may potentially mobilize up to \$100 billion in total investment potential.
- In August 2024, ClearML introduced the AI infrastructure control plane, a universal operating system that offers IT teams and DevOps full control over their AI infrastructure to provide better, self-serve experiences in AI computing and high-performance computing capabilities in clusters.

Some of the Key Companies in the Global AI Infrastructure Market Include:

- Advanced Micro Devices, Inc.
- Alibaba Group
- Baidu Inc.
- Cerebras Systems Inc.
- Cisco Systems, Inc.
- Dell Inc.
- Fujitsu Ltd.
- Graphcore Ltd.
- Groq, Inc.
- Huawei Technologies Co., Ltd.

- Meta Platforms, Inc.
- Micron Technology, Inc
- Microsoft Corp.
- Oracle Corp.
- Qualcomm Technologies, Inc.
- Samsung Electronics Co., Ltd.
- SK Hynix Inc.
- Tesla Inc.

AI Infrastructure Market Segmentation Analysis

Global AI Infrastructure Market by Offering

- Compute
 - o GPU
 - o CPU
 - o TPU
 - o NPU/AI accelerator
 - o Other
- Memory
 - o GDDR
 - o HBM
 - o On-Chip Memory

Global AI Infrastructure Market by Network

- Storage
- Server Software

Global AI Infrastructure Market by Deployment

- On-premises
- Cloud
- Hybrid

Global AI Infrastructure Market by End-User

- Enterprises
 - o Healthcare
 - o BFSI
 - o Automotive
 - o Retail & E-commerce
 - o Media & Entertainment
 - o Others
- Government Organizations
- Cloud Service Providers

Regional Analysis

- North America
 - o United States
 - o Canada
- Europe
 - o UK
 - o Germany
 - o Italy
 - o Spain
 - o France
 - o Rest of Europe
- Asia-Pacific
 - o China
 - o India
 - o Japan
 - o South Korea
 - o ASEAN Economies (Singapore, Thailand, Vietnam, Indonesia, and Other)
 - o Australia and New Zealand
 - o Rest of Asia-Pacific
- Rest of the World
 - o Latin America
 - o Middle East and Africa

Artificial Intelligence (AI) In Asset Management Market: <https://www.omrglobal.com/industry-reports/ai-in-asset-management-market>

Artificial Intelligence (AI) in Cardiology Market: <https://www.omrglobal.com/industry-reports/ai-in-the-cardiology-market>

Artificial Intelligence (AI) in Banking Market: <https://www.omrglobal.com/industry-reports/artificial-intelligence-ai-in-banking-market>

Artificial Intelligence (AI) in Telecommunication Market: <https://www.omrglobal.com/industry-reports/ai-in-telecommunication-market>

Artificial Intelligence (AI) in Construction Market: <https://www.omrglobal.com/industry-reports/artificial-intelligence-ai-in-construction-market>

Anurag Tiwari

Orion Market Research Pvt Ltd

+91 91798 28694

tiwari@omrglobal.com

Visit us on social media:

[Facebook](#)

[X](#)

[LinkedIn](#)

This press release can be viewed online at: <https://www.einpresswire.com/article/804758569>

EIN Presswire's priority is source transparency. We do not allow opaque clients, and our editors try to be careful about weeding out false and misleading content. As a user, if you see something we have missed, please do bring it to our attention. Your help is welcome. EIN Presswire, Everyone's Internet News Presswire™, tries to define some of the boundaries that are reasonable in today's world. Please see our Editorial Guidelines for more information.

© 1995-2025 Newsmatics Inc. All Right Reserved.