+357 96 030922
info@marketpublishers.com

Global AI Chips Supply, Demand and Key Producers, 2026-2032

https://marketpublishers.com/r/GCE850E05AAEEN.html
Date: April 2026
Pages: 190
Price: US$ 4,480.00 (Single User License)
ID: GCE850E05AAEEN

Abstracts

The global AI Chips market size is expected to reach $ 397631 million by 2032, rising at a market growth of 19.5% CAGR during the forecast period (2026-2032).

In 2025, global AI Chips capacity 15,000 k Pcs, sales reached approximately 13,500 k Pcs, with an average market price of around 7,850 USD/Pcs, industrial gross margin 54%.

AI chips have shifted from “a single accelerator” to a system product: in the cloud, GPUs or in-house ASICs are packaged with high-speed interconnects, rack-scale power delivery and liquid cooling to industrialize training and inference as repeatable “AI factories”; on-device, AI chips are embedded into SoCs/CPUs where privacy, power and user experience define differentiation. The market’s center of gravity spans three tracks: NVIDIA-style GPU + full-stack systems, AMD/Intel pushing datacenter accelerators and open software, and hyperscalers (Google/AWS/Microsoft/Meta) scaling in-house ASICs to secure supply and economics; at the edge, Apple and Qualcomm turn NPUs into platform gates.

AI-chip KPIs are moving from peak compute to system efficiency across compute ? memory ? interconnect ? power density. HBM capacity/bandwidth is now decisive (e.g., MI300X discloses 192GB HBM3 and 5.3TB/s; Gaudi 3 up to 128GB HBM and 3.7TB/s; TPU v5p publishes per-chip HBM specs). Low precision (FP8/FP4) and new kernels are the incremental lever for inference and reasoning workloads, while rack-scale fabrics improve realized throughput (NVL72 emphasizes a unified 72-GPU domain and 130TB/s-class fabric; Blackwell Ultra discloses NVLink 5 bandwidth and max topology). Power density and cooling are “chip-level constraints,” with DGX-class systems publishing system power and HBM3e bandwidth; in automotive/robotics, determinism

+357 96 030922
info@marketpublishers.com

and safety isolation become part of what an AI chip is, exemplified by disclosed Thor compute and power envelopes.

Across the supply chain, AI chips are increasingly a systems game: upstream, leading-edge nodes plus advanced packaging (2.5D/CoWoS-class) and HBM supply set the ramp speed; midstream, competition extends from silicon into boards, servers, racks, networking and the software stack (compilers/inference runtimes/collectives) that determines time-to-deploy and stickiness; downstream, hyperscalers, internet platforms, automakers and device OEMs harden requirements into specs. A telling recent move is how rack-scale delivery becomes part of “AI chip capability”: AMD has completed its acquisition of ZT Systems, explicitly combining CPU+GPU+networking with rack-scale systems expertise to match how hyperscalers buy and deploy.

Growth is spilling out along three lines. (1) Datacenter AI chips are shifting from “single-GPU racing” to “AI-factory racing,” with hyperscalers publishing rack/cluster-scale availability (OCI discloses liquid-cooled GB200 NVL72 for very large clusters; SK Group and NVIDIA announce an AI factory exceeding 50,000 GPUs, anchored by manufacturing, digital twins and agents). (2) Geopolitics and compliance have become part of product roadmaps: the U.S. AI diffusion framework and its subsequent rescission—along with continuing guidance on advanced computing ICs—push finer-grained supply, SKU and shipment strategies. (3) On-device AI chips are becoming eligibility gates: Windows writes a 40+ TOPS NPU threshold into Copilot+ PC requirements; Apple discloses M4’s 38 TOPS Neural Engine and ties on-device GenAI to system experience; in vehicles, centralized compute platforms like Thor point to cockpit + AD fusion and higher safety-grade compute as the next leg. Over the next 12–24 months, the most bankable trend is not just “bigger models,” but “more inference, stronger systems, deeper verticalization”: FP4/FP8 inference and communication efficiency become the power-efficiency battlefield; HBM and advanced packaging remain critical supply constraints; sovereign/industry AI factories will expand procurement from chips to racks, power, liquid cooling and ops software; robotics and industrial edge broaden AI-chip demand from vision into multimodal real-time control.

This report studies the global AI Chips production, demand, key manufacturers, and key regions.

This report is a detailed and comprehensive analysis of the world market for AI Chips and provides market size (US$ million) and Year-over-Year (YoY) Growth, considering 2025 as the base year. This report explores demand trends and competition, as well as details the characteristics of AI Chips that contribute to its increasing demand across

+357 96 030922
info@marketpublishers.com

many markets.

Highlights and key features of the study

Global AI Chips total production and demand, 2021-2032, (K Pcs)

Global AI Chips total production value, 2021-2032, (USD Million)

Global AI Chips production by region & country, production, value, CAGR, 2021-2032, (USD Million) & (K Pcs), (based on production site)

Global AI Chips consumption by region & country, CAGR, 2021-2032 & (K Pcs)

U.S. VS China: AI Chips domestic production, consumption, key domestic manufacturers and share

Global AI Chips production by manufacturer, production, price, value and market share 2021-2026, (USD Million) & (K Pcs)

Global AI Chips production by Technical Architecture, production, value, CAGR, 2021-2032, (USD Million) & (K Pcs)

Global AI Chips production by Application, production, value, CAGR, 2021-2032, (USD Million) & (K Pcs)

This report profiles key players in the global AI Chips market based on the following parameters - company overview, production, value, price, gross margin, product portfolio, geographical presence, and key developments. Key companies covered as a part of this study include NVIDIA, AMD, Intel, Google, Microsoft, Amazon, Samsung, Qualcomm, IBM, Apple, etc.

This report also provides key insights about market drivers, restraints, opportunities, new product launches or approvals.

Stakeholders would have ease in decision-making through various strategy matrices used in analyzing the World AI Chips market

Detailed Segmentation:

+357 96 030922
info@marketpublishers.com

Each section contains quantitative market data including market by value (US$ Millions), volume (production, consumption) & (K Pcs) and average price (US$/Pcs) by manufacturer, by Technical Architecture, and by Application. Data is given for the years 2021-2032 by year with 2025 as the base year, 2026 as the estimate year, and 2027-2032 as the forecast year.

Global AI Chips Market, By Region:

United States

China

Europe

Japan

South Korea

ASEAN

India

Rest of World

Global AI Chips Market, Segmentation by Technical Architecture:

GPU

FPGA

ASIC

Others

Global AI Chips Market, Segmentation by Function:

Training Chip

+357 96 030922
info@marketpublishers.com

Inference Chip

Global AI Chips Market, Segmentation by Industry:

Data Center

Automobile

Robot

Consumer Electronics

Medical

Others

Global AI Chips Market, Segmentation by Application:

Cloud Server

Edge and Terminal (Mobile Device)

Companies Profiled:

NVIDIA

AMD

Intel

Google

Microsoft

Amazon

+357 96 030922
info@marketpublishers.com

Samsung

Qualcomm

IBM

Apple

Meta

Cerebras Systems

Groq

Graphcore

Tenstorrent

Hailo

SambaNova

Huawei

Cambricon

Horizon Robotics

Biren

Iluvatar CoreX

Moore Threads

MetaX

Enflame

+357 96 030922
info@marketpublishers.com

Hygon Information Technology

Changsha Jingjia Microelectronics

Kunlunxin (Baidu)

T-Head (Alibaba)

Hexaflake

Key Questions Answered:

1. How big is the global AI Chips market?
2. What is the demand of the global AI Chips market?
3. What is the year over year growth of the global AI Chips market?
4. What is the production and production value of the global AI Chips market?
5. Who are the key producers in the global AI Chips market?
6. What are the growth factors driving the market demand?

+357 96 030922
info@marketpublishers.com

Contents

+357 96 030922
info@marketpublishers.com

3 WORLD MANUFACTURERS COMPETITIVE ANALYSIS

4 UNITED STATES VS CHINA VS REST OF THE WORLD

+357 96 030922
info@marketpublishers.com

+357 96 030922
info@marketpublishers.com

7 MARKET ANALYSIS BY INDUSTRY

8 MARKET ANALYSIS BY APPLICATION

9 COMPANY PROFILES

+357 96 030922
info@marketpublishers.com

+357 96 030922
info@marketpublishers.com

+357 96 030922
info@marketpublishers.com

+357 96 030922
info@marketpublishers.com

+357 96 030922
info@marketpublishers.com

+357 96 030922
info@marketpublishers.com

10 INDUSTRY CHAIN ANALYSIS

+357 96 030922
info@marketpublishers.com

+357 96 030922
info@marketpublishers.com

List Of Tables

LIST OF TABLES

+357 96 030922
info@marketpublishers.com

+357 96 030922
info@marketpublishers.com

+357 96 030922
info@marketpublishers.com

+357 96 030922
info@marketpublishers.com

+357 96 030922
info@marketpublishers.com

+357 96 030922
info@marketpublishers.com

+357 96 030922
info@marketpublishers.com

+357 96 030922
info@marketpublishers.com

+357 96 030922
info@marketpublishers.com

List Of Figures

LIST OF FIGURES

+357 96 030922
info@marketpublishers.com

+357 96 030922
info@marketpublishers.com

+357 96 030922
info@marketpublishers.com

I would like to order

Product name: Global AI Chips Supply, Demand and Key Producers, 2026-2032
Product link: https://marketpublishers.com/r/GCE850E05AAEEN.html
Price: US$ 4,480.00 (Single User License / Electronic Delivery)

If you want to order Corporate License or Hard Copy, please, contact our Customer Service:
info@marketpublishers.com

Payment

To pay by Credit Card (Visa, MasterCard, American Express, PayPal), please, click button on product page https://marketpublishers.com/r/GCE850E05AAEEN.html