+357 96 030922
info@marketpublishers.com

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

https://marketpublishers.com/r/GD72C3542B05EN.html
Date: August 2026
Pages: 216
Price: US$ 4,480.00 (Single User License)
ID: GD72C3542B05EN

Abstracts

The global Large-Model AI Accelerator Chips market size is expected to reach $ 1097725 million by 2032, rising at a market growth of 22.0% CAGR during the forecast period (2026-2032).

Large-Model AI Accelerator Chips are semiconductor processors, accelerator cards, modules and custom compute devices designed to accelerate large language models, multimodal foundation models, generative AI applications, recommender systems and agentic AI workloads. Their core capabilities are built around high-throughput tensor and matrix computation, low-precision arithmetic, high-bandwidth memory access, on-chip and scale-out interconnects, compiler support, runtime software stacks and cluster-level execution. The scope of this study focuses on AI GPUs, AI ASICs, NPUs, TPUs, XPUs, RDUs, LPUs, IPUs and related accelerator modules that can support large-model training, post-training, fine-tuning, inference, token generation and multimodal AI workloads in cloud, data center and high-performance edge environments.

Based on our research, the large-model AI accelerator chip industry has moved beyond a single ?GPU accelerator? framework and is now evolving into a multi-architecture competition among AI GPUs, custom ASICs, TPUs, XPUs, RDUs, LPUs and cloud providers? in-house accelerators. The core of the industry is no longer just peak chip performance; it is a full-stack competition involving advanced process nodes, high-bandwidth memory, advanced packaging, scale-up and scale-out interconnects, system software, compilers, runtime libraries, model compatibility and customer switching costs. NVIDIA remains the structural leader due to its CUDA ecosystem, Blackwell platform and entrenched data center customer base. AMD, Intel, Broadcom, Google, AWS, Microsoft, Meta and Huawei are competing through different approaches, ranging from merchant AI GPUs and Gaudi accelerators to custom ASICs and captive cloud chips.

+357 96 030922
info@marketpublishers.com

For analytical purposes, using ?NPU? alone would materially understate the real industry boundary, so this study adopts ?Large-Model AI Accelerator Chips? as the standard scope.

From the demand-side perspective, frontier model training still requires high-end GPUs and large-scale high-bandwidth clusters, but the incremental growth is increasingly shifting toward inference, real-time token generation, long-context processing, multimodal workloads and agentic AI. Training accelerators are evaluated by FP8/BF16 throughput, HBM capacity, interconnect bandwidth and software maturity, while inference accelerators compete on tokens per watt, tokens per dollar, latency, KV-cache efficiency, model switching speed and rack-level utilization. The expansion of in-house accelerators by Google, AWS, Microsoft and Meta shows that leading cloud customers are increasingly aligning chip design with their own models, serving systems and data-center architectures. The likely market structure is not a simple replacement of GPUs by ASICs, but a workload split in which flexible GPUs remain central for training and heterogeneous development, while custom ASICs gain share in high-volume stable inference.

From the policy and industry-cycle perspective, export controls, domestic substitution, AI data-center capital expenditure, HBM availability and advanced packaging capacity are becoming decisive variables. Chinese vendors still face constraints in leading-edge foundry access, HBM supply, advanced packaging and software ecosystems, but local AI infrastructure demand and policy support are creating a large protected market for domestic accelerators. Over the next three to five years, competition will shift from single-card benchmark claims to system-level total cost of ownership, cluster reliability, model migration cost, inference energy efficiency and supply-chain resilience. The industry remains in a high-growth phase, while substitution risks come from both directions: hyperscalers? in-house ASICs may reduce merchant GPU dependence, and model-level efficiency improvements such as quantization, sparsity, MoE routing and serving-stack optimization may reduce accelerator demand per unit of AI output.

This report studies the global Large-Model AI Accelerator Chips production, demand, key manufacturers, and key regions.

This report is a detailed and comprehensive analysis of the world market for Large-Model AI Accelerator 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 Large-Model AI Accelerator Chips that contribute to its increasing demand across many markets.

+357 96 030922
info@marketpublishers.com

Highlights and key features of the study

Global Large-Model AI Accelerator Chips total production and demand, 2021-2032, (Million Units)
Global Large-Model AI Accelerator Chips total production value, 2021-2032, (USD Million)
Global Large-Model AI Accelerator Chips production by region & country, production, value, CAGR, 2021-2032, (USD Million) & (Million Units), (based on production site)
Global Large-Model AI Accelerator Chips consumption by region & country, CAGR, 2021-2032 & (Million Units)
U.S. VS China: Large-Model AI Accelerator Chips domestic production, consumption, key domestic manufacturers and share
Global Large-Model AI Accelerator Chips production by manufacturer, production, price, value and market share 2021-2026, (USD Million) & (Million Units)
Global Large-Model AI Accelerator Chips production by Type, production, value, CAGR, 2021-2032, (USD Million) & (Million Units)
Global Large-Model AI Accelerator Chips production by Application, production, value, CAGR, 2021-2032, (USD Million) & (Million Units)

This report profiles key players in the global Large-Model AI Accelerator 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 Corporation, Alphabet Inc., Amazon.com, Inc., Broadcom Inc., Microsoft Corporation, Meta Platforms, Inc., Advanced Micro Devices, Inc., Huawei Technologies Co., Ltd., Intel Corporation, Marvell Technology, Inc., 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 Large-Model AI Accelerator Chips market

Detailed Segmentation:

Each section contains quantitative market data including market by value (US$ Millions), volume (production, consumption) & (Million Units) and average price (US$/Unit) by manufacturer, by Type, and by Application. Data is given for the years

+357 96 030922
info@marketpublishers.com

2021-2032 by year with 2025 as the base year, 2026 as the estimate year, and 2027-2032 as the forecast year.

Global Large-Model AI Accelerator Chips Market, By Region:

United States

China

Europe

Japan

South Korea

ASEAN

India

Rest of World

Global Large-Model AI Accelerator Chips Market, Segmentation by Type:

AI GPU

AI ASIC

NPU / XPU

TPU-like Accelerator

Other

Global Large-Model AI Accelerator Chips Market, Segmentation by Numerical Precision:

FP32 / TF32

+357 96 030922
info@marketpublishers.com

BF16 / FP16

FP8

INT8

INT4 / FP4

Other

Global Large-Model AI Accelerator Chips Market, Segmentation by Memory and Packaging:

HBM-based Accelerator

GDDR-based Accelerator

DDR / LPDDR-based Accelerator

Other

Global Large-Model AI Accelerator Chips Market, Segmentation by Interconnect and Scalability:

PCIe-based Scaling

Ethernet-based Scaling

Other

Global Large-Model AI Accelerator Chips Market, Segmentation by Application:

Cloud Computing & AI Infrastructure

Healthcare & Life Sciences

+357 96 030922
info@marketpublishers.com

Energy & Utilities

Consumer Electronics

Other

Companies Profiled:

NVIDIA Corporation

Alphabet Inc.

Amazon.com, Inc.

Broadcom Inc.

Microsoft Corporation

Meta Platforms, Inc.

Advanced Micro Devices, Inc.

Huawei Technologies Co., Ltd.

Intel Corporation

Marvell Technology, Inc.

OpenAI, L.L.C.

Qualcomm Incorporated

Alibaba Group Holding Limited

Cambricon Technologies Corporation Limited

Hygon Information Technology Co., Ltd.

+357 96 030922
info@marketpublishers.com

Cerebras Systems, Inc.

Groq, Inc.

SambaNova Systems, Inc.

d-Matrix Corporation

Tenstorrent Inc.

Rebellions Inc.

FuriosaAI, Inc.

Moore Threads Technology Co., Ltd.

Shanghai Enflame Technology Co., Ltd.

Kunlunxin Technology Co., Ltd.

Shanghai Biren Intelligent Technology Co., Ltd.

MetaX Integrated Circuits Co., Ltd.

Iluvatar CoreX

Hailo Technologies Ltd.

SiMa.ai

Axelera AI B.V.

Etched AI Inc.

Key Questions Answered:
1. How big is the global Large-Model AI Accelerator Chips market?
2. What is the demand of the global Large-Model AI Accelerator Chips market?

+357 96 030922
info@marketpublishers.com

3. What is the year over year growth of the global Large-Model AI Accelerator Chips market?
4. What is the production and production value of the global Large-Model AI Accelerator Chips market?
5. Who are the key producers in the global Large-Model AI Accelerator 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

5 MARKET ANALYSIS BY TYPE

+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

+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

+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 Large-Model AI Accelerator Chips Supply, Demand and Key Producers, 2026-2032

Product link: https://marketpublishers.com/r/GD72C3542B05EN.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/GD72C3542B05EN.html