AI Chip

ICV TAnK
Technology Advisory
& Knowledgebase

Market Research Report
2023

Introduction

ICV TAnK
Technology Advisory
& Knowledgebase

This report mainly introduces artificial intelligence chips and their classification, market size, and applications.

Artificial intelligence chips, also known as AI accelerators or computing cards, are computing chips specialized in artificial intelligence algorithms.There are currently two development paths for artificial intelligence chips. One is a traditional computing architecture represented by GPU, FPGA, ASIC (TPU, BPU, etc.); Another approach is to subvert the classic von Neumann computing architecture and use brain like neural structures to enhance computing power.

According to the deployment location, AI chips can be divided into cloud AI chips, edge AI chips, and terminal AI chips. In the future, With more and more intelligent devices such as smart cars, smartwatches, smart appliances, VR/AR, etc. entering people's lives, the amount of data collected by edges and terminals is increasing exponentially, the integration of edge cloud computing power layout solutions will become mainstream.

ICV TAnK
Technology Advisory
& Knowledgebase

AI Chip Development History

AI Chip Development History

ICV TAnK
Technology Advisory
& Knowledgebase

- AI chips are the underlying cornerstone of AI development.
- NVIDIA invented the GPU as early as 1999, but it was not until 2009 that Stanford University published a paper describing how to use the computing power of modern GPUs to far exceed that of multi-core CPUs (more than 70 times) to shorten AI training time from weeks to hours.

1950s-1960s	1970s	1980s	1990s	2000s	2010s	2020s	
1958	1971	1982	1991	2006	2011	2021	2023
Invention of the Integrated Circuit	Intel 4004, the first microprocessor	Introduction of the first RISC processors	Emergence of Digital Signal Processors (DSPs)	Multi-core processors become mainstream	Rise of AI-specific chips like Google's TPU	Development of advanced neural network processors	Integration of AI capabilities in mainstream processors

The Development Path Of AI Chip

ICV TAnK
Technology Advisory
& Knowledgebase

TYPES	DEGREE OF CUSTOMIZATION	ADVANTAGE	DISADVANTAGE
GPU	General	• Strong computing power • Product mature	• Low efficiency • Programming is difficult
FPGA	Semi customization	• High average performance • Low power consumption • High flexibility	• Peak computing power is weak • Programming language is difficult
ASIC	Fully customized	• Strong average performance • Low power consumption • Small size	• Not editable • Long development time • High technical risk

There are currently two development paths for artificial intelligence chips. One is a traditional computing architecture represented by GPU, FPGA, ASIC (TPU, BPU, etc.)

Another approach is to subvert the classic von Neumann computing architecture and use brain like neural structures to enhance computing power.

ICV TAnK
Technology Advisory
& Knowledgebase

AI Chip Classification

Dimension 1: Undertake Tasks

ICV TAnK
Technology Advisory
& Knowledgebase

According to the different tasks undertaken, AI chips are divided into training chips and inference chips.

- **Training chip**: A complex neural network model needs to be trained through big data and trained systematically using labeled data to adapt to specific functions, emphasizing absolute computing power.
- **Inference chip**: Utilizing trained neural network models, using new data for inference prediction, focusing on comprehensive indicators such as unit energy consumption, time delay, cost, etc.

AI Training And Inference

If it is the image recognition AI of cats and dogs, then.....

Training ---AI views many images and grasps the characteristics of dogs and cats.

Inference---When AI sees an image, it can determine whether it is a cat or a dog.

Training Chip VS Inference Chip

TYPE	FEATURES
Training Chip	• Computing chips used to build neural network models • Training requires high computing performance (high throughput) and low power consumption • More emphasis is placed on absolute computing power
Inference Chip	• Mainly use trained neural network models for inference and prediction • Pay more attention to low latency and low power consumption • Requires relatively low computing power

Dimension 2: Deployment Location

ICV TAnK
Technology Advisory
& Knowledgebase

Cloud

The cloud AI chips are responsible for processing and managing data and tasks of cloud computing, mainly implementing AI training and inference processes.

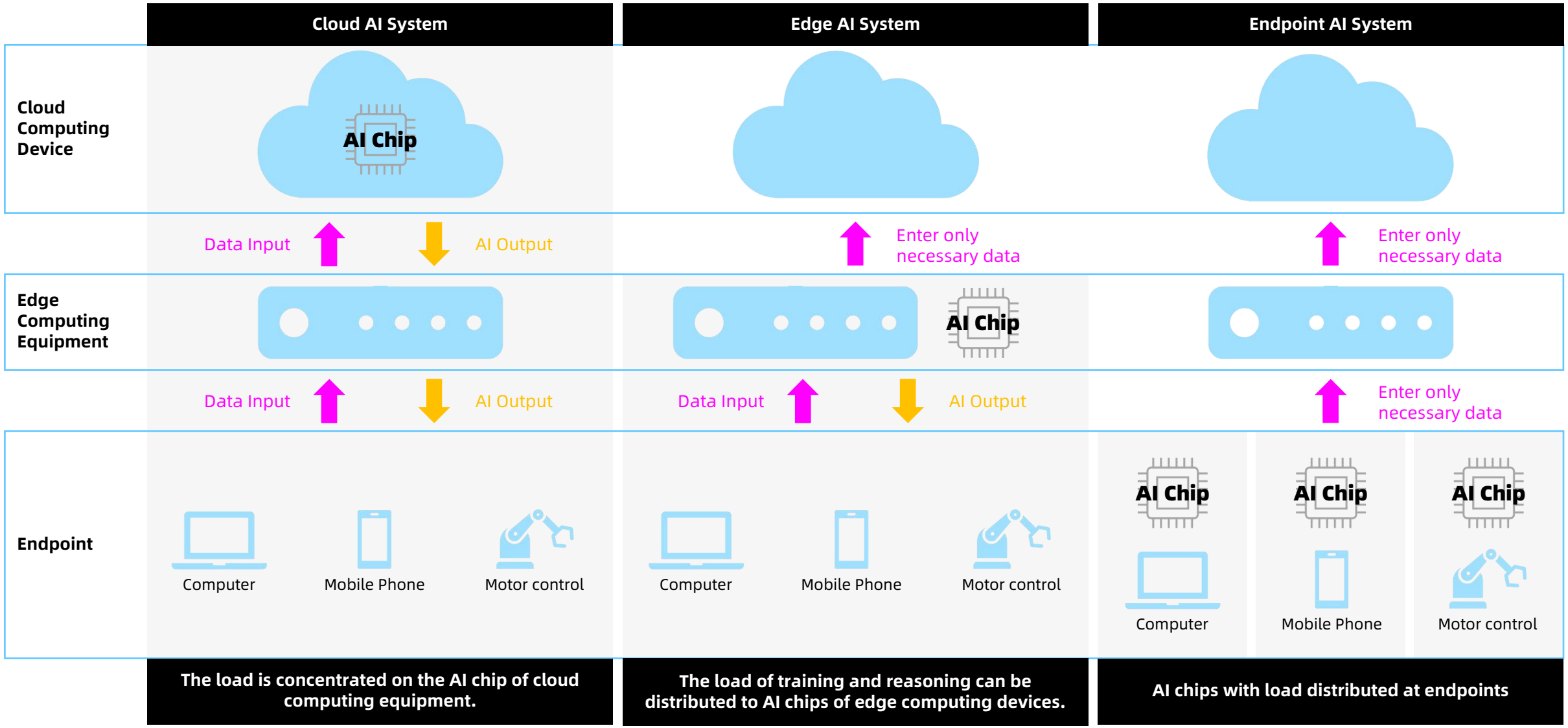
Edge

The edge AI chips can achieve local processing and storage of data without the need for internet connection, mainly achieving AI inference.

Terminal

The terminal AI chips is mainly responsible for executing user instructions and operations, as well as communicating and interacting with the cloud and edge, mainly realizing the artificial intelligence inference process.

INDEX	CLOUD AI CHIP	EDGE AI CHIP	TERMINAL AI CHIP
Processing capacity	Have the strongest processing power	Compromise in processing power compared to cloud chips	In terms of processing ability, it is the weakest among the three
Delay	High latency	Reduces latency	Extremely low latency
Security	More stringent security measures are needed	Higher data security	Maximizing privacy protection
Application	The application of large-scale data processing and complex AI models	Applications that require quick response and partial autonomous decision-making，Such as autonomous vehicle	Suitable for small, personalized devices



ICV TAnK
Technology Advisory
& Knowledgebase

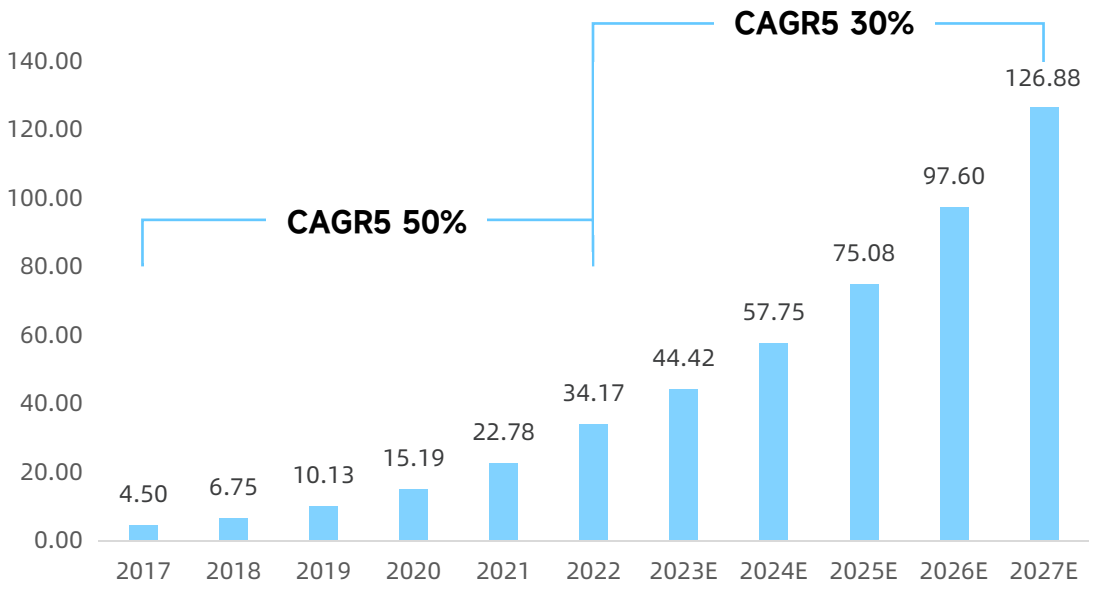
AI Chip Market Size

The Global AI Chips Market Is Growing Rapidly

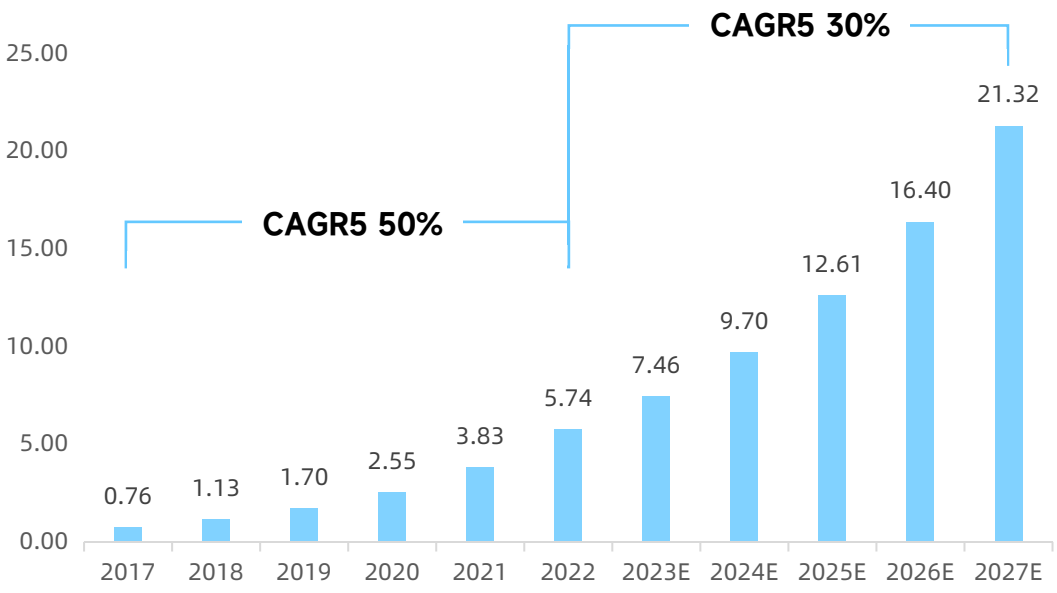
As the AIGC industry enters a period of rapid development, the performance of AI large models has been continuously improved. As the core of computing power, AI chips have been able to grow rapidly driven by the development of the AIGC industry. The global AI chip market has grown from 4.5 billion dollars in 2017 to 34.17 billion dollars in 2022.

From the perspective of the Chinese market, the size of the AI chip market in China has increased from 0.76 billion dollars in 2017 to 5.74 billion dollars in 2022, with a compound growth rate of 50%. It is expected to occupy a global market share of 16.80% by 2027.

Global AI Chip Market （$,Billion）



China AI Chip Market （$,Billion）

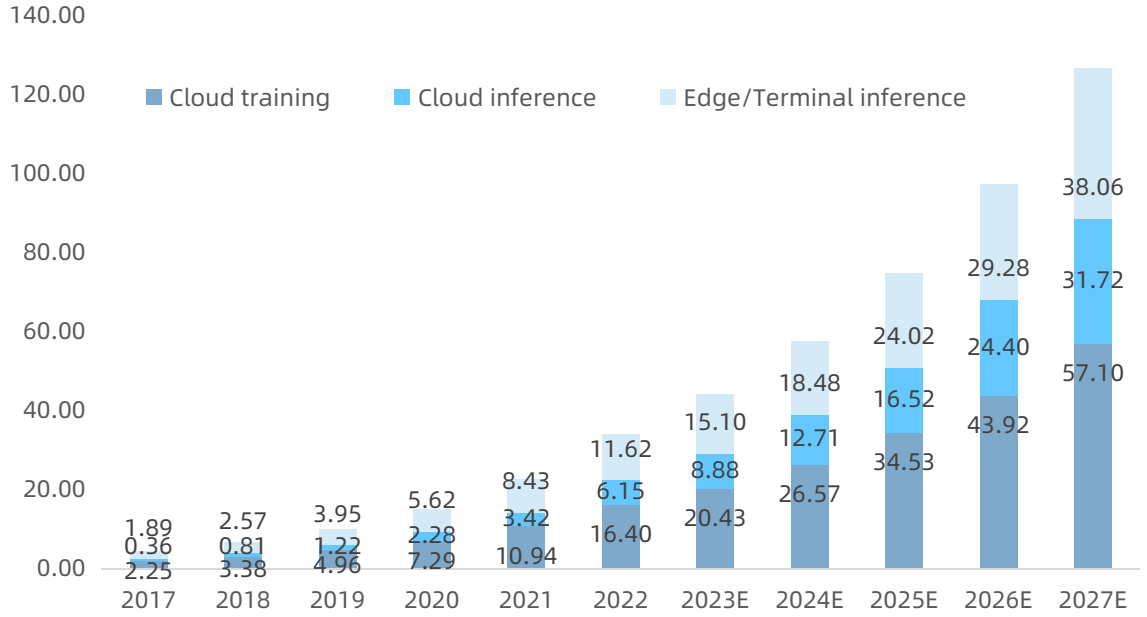


Cloud Training Chips Account For The Main Market

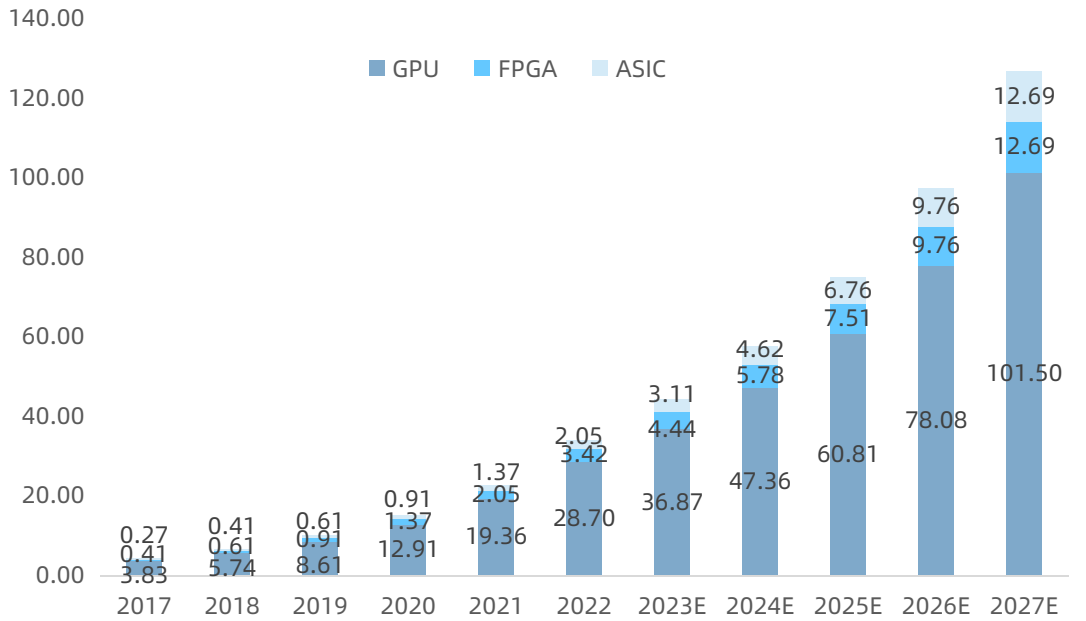
Driven by the continuous application and penetration of AI big models, cloud computing, and data centers, AI chips have developed rapidly. Especially for cloud training chips, the global market size of cloud training chips reached 16.40 billion dollars in 2022, accounting for approximately 48% of the entire AI chip market.

From another perspective, AI chips mainly include graphics processors (GPUs), field programmable gate arrays (FPGAs), and specialized integrated circuits (ASICs). Among them, GPUs hold the main market share in AI chips, account for 84% of AI chip products in 2022, with a market size of 28.70 billion dollars. It is expected that this size will reach 101.50 billion dollars by 2027.

Global AI Chip Segmentation Market （$,Billion）



Market Of AI Chip With Different Architectures （$,Billion）



iCV TAnK
Technology Advisory
& Knowledgebase

AI Chip Market Players

Global AI chip industry map

ICV TAnK
Technology Advisory
& Knowledgebase

Cloud Chip

NVIDIA | AMD | intel | Google

Microsoft | Cambricon 寒武纪

HUAWEI | 昆仑芯 | 壁仞科技 BIREN TECHNOLOGY | Enflame 燧原科技

摩尔线程 MOORE THREADS | METAX 沐曦 | 天数智芯 | 平头哥

Edge/Terminal AI Chip

NVIDIA | intel | arm

Qualcomm | Google | SAMSUNG | 地平线 Horizon Robotics

黑芝麻智能 BLACK SESAME TECHNOLOGIES | intellifusion 云天励飞

LYNXI | ALLWINNER | Rockchip 瑞芯微电子

Nvidia Has Absolute Advantage Of Cloud Chips

ICV TAnK
Technology Advisory
& Knowledgebase

In the field of cloud training chips, Nvidia has an absolute advantage . Nvidia's GPU+CUDA computing platform is currently the most mature AI cloud training solution, along with international vendors such as Google, Intel, AMD, and Xilinx. Due to their relatively late entry into this field, Chinese manufacturers are still primarily start-up companies and have not yet formed an influential cloud based training chip ecosystem.

Compared with cloud training chips, cloud inference chips consider more comprehensive factors, including unit power consumption, latency, cost, etc. FPGA/ASIC chips have more outstanding performance and application advantages.In the cloud inference chip market, international manufacturers such as Nvidia, Google,, and Intel dominate the market, while Chinese companies such as Cambrian, BitContinent, and Huawei are also actively expanding their presence.

Edge AI Chips Have Development Prospects

ICV TAnK
Technology Advisory
& Knowledgebase

At present, the edge/terminal inference chips are each in an array, and there is no chip manufacturer with an absolute dominant position.

Among them, the AI terminal inference chip market in the security field is relatively stable, with Nvidia and Mobileye as industry leaders, and HiSilicon and Anba forming strong competition with them.

The mobile phone market is mainly dominated by manufacturers of original control chips from Qualcomm, Huawei, and Apple;NVIDIA, TI, Renesas, and NXP are the main participants in the field of autonomous driving.

	IoT Scenarios	Mobile Internet	Intelligent Security	Autonomous Driving
Task	• Image detection • Video detection • speech recognition	• Photo -scene recognition • AR application • Voice Assistant	• Image detection • Video detection	• Data fusion • Planning • Image semantic segmentation
Energy Consumption	Deploy on-site power supply for device access and deployment	Consumer grade polymer lithium batteries	Deploy on-site power supply for device access and deployment	Power grade hard shell lithium battery
Reliability	High	Medium	High	Extremely high
Manufacturers	Google,NVIDIA	Apple, Samsung,ARM	Ingenic, Intell if usion	NVIDIA,Intel

ICV TAnK
Technology Advisory
& Knowledgebase

AI Chip Development Trends

AI Chips Are Moving Towards Edge AI Chips

ICV TAnK
Technology Advisory
& Knowledgebase

With more and more intelligent devices such as smart cars, smartwatches, smart appliances, VR/AR, etc， the amount of data collected by edges and terminals is increasing exponentially, and there are higher requirements for real-time response and low latency. More intelligent data processing will be completed on the edge side.Compared to cloud AI chips, edge AI chips can significantly reduce dependence on cloud computing resources, reduce network bandwidth consumption, reduce data transmission latency, improve processing efficiency, and optimize user experience.

However, diversified application scenarios also bring some problems to edge AI, such as the need to customize different algorithms for different application scenarios, and the requirements for chip computing power and power consumption are also different. There are also situations where computing power, algorithms, and applications are fragmented.

In the future, the integration of edge cloud computing power layout solutions will become mainstream, not only achieving optimization of algorithm structure, but also fundamentally empowering various edge applications to provide more complete solutions.

About ICV TAnK

At ICV we are passionately curious about New Technology, and we strive to deliver the most robust market data and insights, to help our customers make the right strategic decisions.

We bring together the deepest intelligence across the widest set of capital-intensive industries and markets. By connecting data across variables, our analysts and industry specialists present our customers with a richer, highly integrated view of their world.

That is the benefit of The New Intelligence. We are able to isolate cause and effect, risk and opportunity in new ways that empower our customers to make well-informed decisions with greater confidence.

AI Chip

ICV TAnK
Technology Advisory
& Knowledgebase

Our offices

To find an office address, phone or fax number, please select a location from the list below. View our global offices.

Canad

5250 Fairwind Dr. Mississauga,
Ontario,
L5R 3H4,
Canada

Singapore

101 Upper Cross Street,
#04-17,
People’s Park Centre,
Singapore

More research report in www.icvtank.com