The information and statistics set out in this section and other sections of this document were extracted from a report prepared by CIC under our commission, various official government publications and other publicly available sources. We engaged CIC to prepare an independent industry report, or the CIC Report, in connection with the **[REDACTED]**. *We believe that the sources of this information are appropriate sources for such information and have taken reasonable care in extracting and reproducing such information. We have no reason to believe that such information is false or misleading or that any fact has been omitted that would render such information false or misleading. The information from official government sources has not been independently verified by us, the Joint Sponsors, the* **[REDACTED]**, *the* **[REDACTED]** *or any other party involved in the* **[REDACTED]** *and no representation is given as to its accuracy.*

SOURCE OF INFORMATION

We engaged CIC, an independent market research and consulting company that provides industry consulting services, commercial due diligence, and strategic consulting, to conduct detailed research on and analysis of China's AI inference chip-related products and services industry and smart device industry. We have agreed to pay a fee of RMB869,200 (inclusive of tax) to CIC in connection with the preparation of the CIC Report, which we believe to be consistent with market rates. We are of the view that the payment of such fee does not impair the fairness of the conclusions drawn in the CIC Report.

In preparing the CIC Report, CIC conducted both primary and secondary research, and gathered knowledge, statistics, information, and insights on industry trends within the target research markets. The primary research involved interviews with key industry experts and leading industry participants. The secondary research consisted of analyzing data from various publicly available sources, such as the National Bureau of Statistics.

The CIC Report was compiled based on the following assumptions: (1) the overall political, economic and social environment in China is expected to remain stable during the forecast period; (2) related key industry drivers are likely to propel continued growth in China's AI inference chip-related products and services industry and smart device industry throughout the forecast period, including favorable policies; and (3) there will be no extreme force majeure or unforeseen industry regulations in which the market may be affected in either a dramatic or fundamental way during the forecast period.

DIRECTORS' CONFIRMATION

After making reasonable inquiries, our Directors confirm that, to the best of their knowledge, there has been no detrimental change in the market information demonstrated in the CIC Report since the date of the report that may qualify, contradict or have an impact on the information in this document.

RAPID GROWTH IN AI INFERENCE DRIVEN BY LARGE-SCALE MODEL APPLICATIONS

The rapid advancement of AI, big data, and cloud computing is driving a sweeping wave of intelligent transformation, fundamentally reshaping production models and competitive landscapes across industries. AI is increasingly solving real-world problems in sectors such as transportation, internet search, and manufacturing, thereby accelerating the intelligent transformation of society.

The parameter size of large-scale models continues to grow, with the cost of training models containing hundreds of billions of parameters surpassing the tens of millions of dollars threshold. Marginal returns are diminishing. Meanwhile, the disruptive decline in inference costs has sparked an explosive surge in inference-based applications. The emergence of domestic open-source models such as DeepSeek has further lowered development barriers by enabling scenario-specific fine-tuning without full training cycles. In addition, progress in domestic chip alternatives is creating a more reliable hardware foundation for large-scale deployment.

AI has progressed from iterative algorithmic improvements to robust advancements in computing infrastructure, fueling innovation in foundational hardware. Within this evolution, AI chips—serving as the core hardware enabling intelligent computing—are emerging as a critical component of the value chain. AI development is also redefining hardware products. Large models are now catalyzing two parallel hardware evolution pathways, including AI-native products, which are entirely new product categories born from rapid AI advancements, and AI-empowered products driven by the enhancement of existing products through AI capability integration.

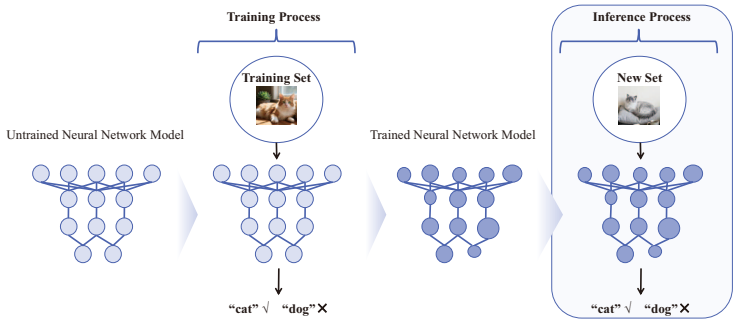
Definition and Classification of AI Chips

AI chips have become indispensable hardware for supporting large-scale computing tasks, playing critical roles in both training and inference—the two fundamental phases of AI model development and deployment. *Training* involves developing AI models using vast datasets and machine learning algorithms. As the cornerstone of AI development, this phase focuses on designing model architectures and optimizing parameters. Its computational demands are driven by the exponential growth in model complexity and training scale. *Inference*, in contrast, applies trained models to new data, enabling tasks such as image, speech, text recognition, classification, and prediction. Its performance hinges on concurrency handling and response speed, tailored to specific deployment scenarios.

Given the stark differences in computational intensity, processing frequency, and deployment environments between these phases, specialized AI chips are required for each: *Training chips* excel at building large-scale AI models, prioritizing high computational precision, massive memory bandwidth, and high throughput. They are primarily deployed in cloud data centers to handle intensive workloads. *Inference chips*, optimized for efficiency, emphasize low latency, cost-effectiveness, power efficiency, and scalability. They power diverse environments—from cloud platforms to edge and terminals—meeting real-time demands across applications.

AI inference refers to the process whereby a trained model utilizes its learned parameters to analyze new, previously unseen data and generate classification results or predictive outcomes. It represents the critical stage in which an AI model transitions from "learning" to "application."

Training and Inference Process



Source: CIC Report

Note: Labeled data from the training set are fed into an untrained model, for example by using an image labeled as "cat" as a training sample. The model iteratively adjusts its internal parameters based on the discrepancy between its predictions and the true labels, gradually learning relevant features and developing stable recognition capability, ultimately becoming a trained model. After training is completed, the model enters the inference phase, during which it receives an image from the new set that was not included in the training set, and applies the features learned during training to analyze the input and generate a prediction result.

Architecture Evolution of AI Inference Chips

With the rapid popularization of AI applications, to meet the market demand for highly power efficiency AI inference systems with extreme low cost and extreme high performance, the architecture of AI inference chips has undergone a fundamental transformation, shifting from CPU-based general computing and GPU-based general parallel computing to NPU-based AI inference-optimized parallel computing. In the early stages of AI, AI training and AI inference relied on the CPU-based IT infrastructure which were designed for general computing. However, the limitations of CPUs in parallel processing capability and power efficiency prompted the industry to adopt GPUs with general parallel computing capabilities. Currently, the AI training workloads are running on GPUs. Nevertheless, AI training and AI inference are two distinct tasks with significant differences. Although GPUs offer superior parallel computing capabilities, the architecture of GPUs, which is oriented toward graphics processing, is not the optimal choice for tensor operations, such as large-scale matrix operations, especially for AI inference that only focuses on the forward computing of neural networks. As deep learning models used in AI applications grew more complex, exposing GPU shortcomings in power consumption and computational density, the industry pivoted to purpose-built neural processing units (NPUs).

NPUs represent a quantum leap in AI inference acceleration, featuring native support for tensor operations, low-precision quantization, and large-scale parallel processing. These architecture-specialized chips achieve unprecedented improvements in four critical dimensions: (1) reduced latency (enabling real-time inference), (2) enhanced power efficiency (maximizing throughput per watt and expanding deployment to edge devices and terminals), (3) increased computational density (chips can accommodate stronger computing performance) and (4) optimized cost efficiency (facilitating broader and more sustainable adoption across diverse application scenarios). Through hardware-software co-design, NPUs achieve algorithmic-aware optimization, transforming AI inference from a functional capability to a scalable, cost-effective, and production-ready solution.

This architectural evolution from general-purpose CPUs to semi-specialized GPUs, and then to fusion-architecture GPNPUs or specialized-architecture NPUs/ASICs optimized for AI inference workloads mirrors the maturation of AI deployment. The trend of scenario-specific optimization will intensify, with future chips and systems increasingly tailored to the unique computational patterns of vertical applications. This hardware specialization, coupled with algorithmic refinement, is creating a virtuous cycle that propels AI from laboratories to real-world impact.

Main Deployment Locations and Classification of AI Inference Chips

AI inference chips can be categorized into three types based on deployment location: cloud-based inference chips, edge inference chips, and terminal inference chips.

- *Cloud-based inference chips* are deployed in data centers and public cloud servers, offering 100 to over 1,000 TOPS of computing power for high-concurrency tasks and large model inference. While they provide exceptional processing capabilities and elastic scalability, their high power consumption presents significant energy management challenges.

- *Edge inference chips* operate between cloud servers and terminals, typically delivering 10-300 TOPS of computing power for applications such as edge gateways and robots. These chips prioritize low latency, high reliability, and balanced power efficiency, enabling local data processing that reduces cloud dependency and bandwidth requirements. Their middle computing performance makes them ideal for scenarios demanding rapid response times and stable operation near data sources. They are also capable of running compressed AI models, supporting diverse edge intelligence tasks.

- *Terminal inference chips* are integrated directly into lightweight commercial products such as smartphones (1-50 TOPS+) and wearables (0.1-0.5 TOPS), emphasizing ultra-low power consumption and compact designs. These chips enable millisecond-level local inference, enhancing real-time responsiveness while optimizing battery life. As lightweight large-scale models are increasingly being deployed at the terminal level, terminal inference chips play a critical role in supporting applications such as intelligent voice assistants, real-time translation, and image recognition. By processing data on terminals, they significantly improve user experience in mobile applications while minimizing cloud transmission needs.

Technical Evolution of AI Inference Chip-Related Products and Services

The rapid expansion of AI applications across industries is driving three critical requirements for inference chips: greater cost-efficiency, greater power efficiency, and enhanced scenario adaptability. These demands are accelerating innovation across four key technological frontiers.

- *Co-optimization of algorithms and chip architectures*. Advanced co-optimization approaches now integrate the entire development chain, from algorithm analysis to toolchain development, creating hardware specifically tuned for target workloads. This synergy delivers faster, more accurate inference while maximizing resource efficiency, particularly in area- and power-constrained environments.

- *Neural network efficiency and power consumption optimization.* Edge and endpoint deployments require perfect balance between throughput and power efficiency. Cutting-edge techniques optimize MAC (multiply-accumulate) unit structures, implement sparse computation, and streamline memory access paths—maintaining model accuracy while dramatically reducing energy consumption.

- *Processing-in-memory ("PIM") technology.* Traditional chip architectures separate computation and memory, creating energy-intensive data transfers that slow AI processing. Near-memory computing partially addresses this by positioning compute units closer to memory. PIM technology embeds computation directly within memory cells, enabling data processing without movement. This breakthrough dramatically boosts efficiency and speed, making PIM architecture essential for future AI chips.

- *Chiplet-based heterogeneous integration.* AI's complexity outpaces monolithic chip capabilities. Chiplet technology decomposes systems into modular NPU, memory, I/O interfaces, and memory blocks, enabling mix-and-match integration on advanced packaging substrates. This paradigm shift delivers three key advantages, including better performance through optimized IP reuse, improved yields via smaller dies, and faster time-to-market through parallel development. Chiplet integration is emerging as a key approach in the next generation of high-performance AI chip design.

Application Scenarios of AI Inference Chip-Related Products and Services

AI inference chip-related products and services serve as essential infrastructure across enterprise, consumer, and industrial applications. In enterprise settings, these chip-related products and services power AI inference servers, edge gateways, and robots for internet companies, AI companies, telecom operators, and research institutions. They provide the computational muscle needed for real-time, large-scale inference tasks that drive business operations and innovation.

For consumer applications, the technology manifests through on-device, cloud-based or hybrid implementations. Wearables incorporate specialized chips emphasizing ultra-low power consumption and compact designs to deliver responsive, privacy-conscious AI experiences. Meanwhile, cloud deployments focus on supporting high concurrency and power efficient processing to handle dynamic user demands across services like smart devices for children, learning devices and smart home devices.

Industrial applications leverage these chip-related products and services through tailored edge and cloud solutions designed for specific vertical needs. From smart city infrastructure to intelligent transportation systems, the technology enables localized processing, custom functionality, and real-time responsiveness. The adaptability of chip-related products and services' across these diverse scenarios underscores their transformative potential throughout the AI value chain.

Demand for China's full-scenario AI inference chip-related products and services

China's full-scenario AI inference chip-related products and services industry encompasses a broad range of hardware and software solutions designed to support AI inference across diverse application scenarios, including enterprise, consumer, and industrial applications. The industry enables a comprehensive full-scenario layout that ensures AI capabilities are seamlessly embedded into all levels of the digital ecosystem.

AI inference chip-related products and services, together with smart devices, form a tightly integrated system that enables efficient, adaptive, and intelligent computing. By working in close synergy, they support comprehensive deployment across cloud, edge, and terminals, ensuring intelligent processing capabilities are available across full scenarios. AI inference chip-related products and services, which serve as the computational backbone, are increasingly embedded in smart devices to meet the growing demand for real-time, on-device AI across diverse application scenarios. Smart devices actively drive AI innovation by collecting valuable real-world data, including visual, audio, semantic, and behavioral data, and providing crucial performance feedback. Moreover, operational metrics and identified performance bottlenecks from deployed devices offer critical insights for the co-optimization of algorithms and chip architectures, creating a continuous improvement cycle where practical applications inform both software and hardware advancements.

The development of full-scenario ecosystems not only drives the evolution of AI inference chip but also facilitates the expansion into downstream applications such as smart devices, offering significant advantages across the entire value chain. AI application scenarios are steadily expanding from industrial manufacturing and smart cities to consumer electronics, enhancing smart devices with stronger data processing and autonomous decision-making capabilities. Full-scenario deployment not only increases the practical value of AI inference chips and smart devices but also unlocks broader market potential for the industry. Against this backdrop, companies with full-scenario capabilities are positioned to gain a stronger competitive edge.

Amid the ongoing expansion of full-scenario AI capabilities, the architectural evolution of NPUs is not limited to advancements in chip design but has also driven transformative changes in smart devices. Smart devices are highlighted separately here, as they carry inherent brand value in addition to serving as platforms for computation, data generation, and intelligent user interaction. The following content will be divided into two parts for analysis: AI inference chip-related products and services and smart devices, as these two components differ in their respective value propositions.

OVERVIEW OF CHINA'S AI INFERENCE CHIP-RELATED PRODUCTS AND SERVICES INDUSTRY

Definition of AI Inference Chip-Related Products and Services

The AI inference chip-related products and services combine specialized hardware and software components to deliver real-time AI processing capabilities. In addition, they are often accompanied by computing power services, further addressing diverse inference workloads and forming a more comprehensive, end-to-end solution. At its core, the hardware features power efficient chip architectures paired with optimized deployment frameworks. The software stack forms the intelligent backbone, comprising key elements like algorithm models, compiler toolchains, instruction sets, and processor architectures, all working in concert to maximize hardware potential through scenario-specific

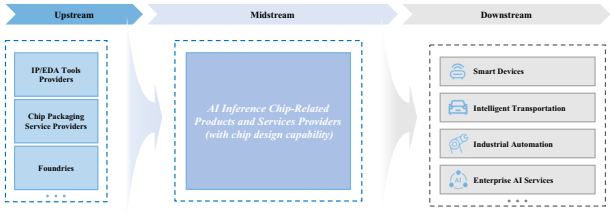
optimizations. This holistic approach enables developers to fully leverage chip capabilities while dramatically reducing deployment complexity, bridging the gap between theoretical AI models and practical, high-performance implementations.

Value Chain Analysis of AI Inference Chip-Related Products and Services

The industry value chain for AI inference chip-related products and services comprises three core segments. The upstream segment includes IP providers, which design and license reusable semiconductor intellectual property blocks such as processor cores, memory controllers, and interface standards; EDA tool vendors, which supply electronic design automation software for chip design, verification, and testing; as well as chip packaging and testing companies and foundries, all of which offer the essential technological foundations and manufacturing capabilities. The midstream segment is centered around AI inference chip-related product and service providers, who leverage their innovation capabilities and accumulated technical expertise to carry out circuit design and verification, while carefully balancing performance, power consumption, and cost. The downstream segment involves the deployment of AI inference solutions across a wide range of application scenarios. The diversity of downstream demand drives chip design companies to build extensive product portfolios, enabling them to meet varied customer needs through flexible product combinations.

Our products and services in enterprise- and industrial-class scenarios operate primarily within the midstream and downstream of the industry value chain.

Value Chain of AI Inference Chip-Related Products and Services



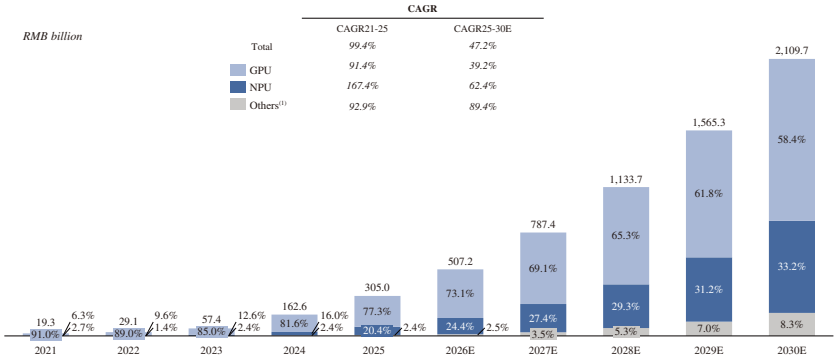
Source: CIC Report

Market Size of AI Inference Chip-Related Products and Services Industry in China

China has witnessed a rapid growth in the demand for AI chips. The market size of AI chips-related products and services market in China expanded from RMB28.4 billion in 2021 to RMB427.3 billion in 2025, at a CAGR of 97.0%, and is expected to reach RMB2,637.5 billion by 2030, at a CAGR of 43.9% from 2025 to 2030.

The AI inference chip-related products and services industry in China is experiencing rapid growth. The market size expanded from RMB19.3 billion in 2021 to RMB305.0 billion in 2025, at a CAGR of 99.4%, and is expected to reach RMB2,109.7 billion by 2030, at a CAGR of 47.2% from 2025 to 2030. The NPU-powered market size grew from RMB1.2 billion in 2021 to RMB62.1 billion in 2025, with a CAGR of 167.4%, and is projected to climb to RMB701.2 billion by 2030, reflecting a CAGR of 62.4% during 2025 to 2030.

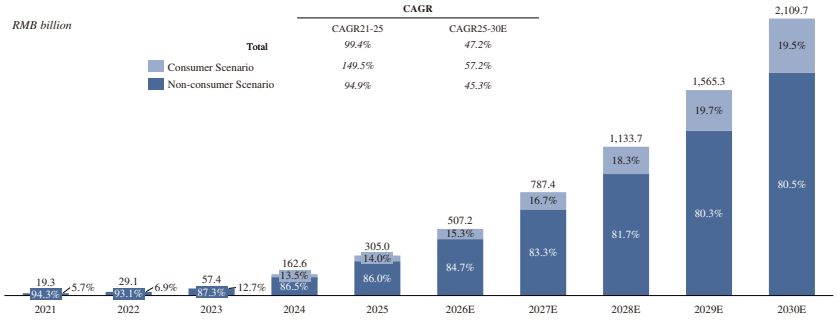
Market Size of China's AI Inference Chip-Related Products and Services Industry, in terms of Revenue, 2021-2030E



Source: IDC, CIC Report

(1) "Others" means other AI-capable hardware components that are not strictly categorized as GPUs or NPUs, but are also used in AI inference scenarios

Market Size of China's AI Inference Chip-Related Products and Services Industry, in terms of Revenue, by Downstream Industry, 2021–2030E



Source: IDC, China Insights Consultancy

The consumer scenario is the fastest-growing segment in China's AI inference chip-related products and services market, recording a CAGR of 149.5% between 2021 and 2025, compared with 94.9% for the non-consumer scenario, which mainly refers to enterprise and industry scenarios. Nevertheless, both segments are expected to maintain strong growth momentum during 2025-2030, with the consumer scenario projected to further expand its share.

Key Drivers and Trends for China's AI Inference Chip-Related Products and Services Industry

Key growth drivers and trends of China's the AI inference chip-related products and services industry include the following.

- *Surging demand for AI inference.* The explosive growth of large AI models has become a primary catalyst for inference demand. Applications like AI-generated content, virtual assistants, and digital humans are generating exponentially increasing inference requests.

- *Enterprise digital transformation.* Businesses are driving demand through widespread adoption of AI-powered automation and analytics. Enterprises are demonstrating strong demand for both public cloud and private AI deployments. The demand, combined with growing pressure to optimize operational efficiency, has made cost-effective AI inference chip-related products and services a strategic priority for corporate IT infrastructure.

- *Consumer electronics revolution.* The consumer market presents unique challenges for inference chip deployment. Smart wearables like earbuds and watches require chips that combine high performance with ultra-low power consumption to overcome strict size, battery, and thermal constraints. Rising consumer expectations for real-time, personalized AI experiences with strong privacy protections are further accelerating adoption of efficient on-device inference solutions.

- *Government-led infrastructure demand.* At the industrial level, national policy initiatives are actively driving the development of next-generation cities characterized by digitalization, interconnected networks, and intelligent systems. For example, the General Office of the State Council has proposed the development of key projects such as comprehensive urban operation management service platforms and City Information Modeling ("CIM") platforms. These large-scale projects require massive computing infrastructure, making operational cost control a critical concern. As a result, AI inference chip-related products and services must deliver exceptional performance-to-power ratios to meet the demands of urban-scale deployments and intelligent industrial parks.

- *Continuous technological advancements.* The supply side is responding with rapid innovation in chip architectures and manufacturing processes. Concurrent improvements in supporting toolchains and model compression techniques are significantly reducing deployment barriers. These technological iterations are enhancing every aspect of inference chips, from power efficiency to computational density.

- *Compelling economic advantages.* Inference chips hold distinct advantages over training solutions in cost efficiency and deployment flexibility. As production scales up, supply chains optimize, and flexible and shared resource allocation is achieved, the cost per unit of computing power continues to decline. This superior cost-performance ratio makes AI inference chip-related products and services increasingly attractive for commercial deployment across all market segments.

- *Expansion of application scenarios.* AI inference chip-related products and services are expanding into new vertical applications, moving beyond traditional uses to power advanced domains like autonomous driving, robots and AI agents. These cutting-edge applications demand significantly higher performance—including sub-millisecond latency, fail-safe reliability, and autonomous decision-making capabilities. As inference solutions become deeply integrated into specialized industry workflows, they are evolving from basic functional tools to optimized, high-efficiency systems that enable widespread commercial deployment.

- *Rising NPU penetration.* The integration of NPUs, which are better suited for AI inference, into a wide range of terminals has become a clear industry trend. This trend is reflected in the steadily rising penetration of NPUs in AI inference chips, increasing from 6.3% in 2021 to 20.4% in 2025, and projected to reach 33.2% by 2030. NPU-powered hardware is rapidly proliferating, reshaping deployment pathways and accelerating the expansion of the AI inference chip ecosystem.

OVERVIEW OF CHINA'S SMART DEVICE INDUSTRY

Definition of Smart Devices

Smart devices serve as critical infrastructure that bridges cloud platforms, edge nodes, and end-user environments, and with substantial scale. These AI-enhanced terminals now integrate advanced perception, real-time processing, and autonomous decision-making capabilities, transforming from simple control units into sophisticated cognitive systems. This architectural shift enables devices to not just collect data but understand contexts and initiate appropriate actions, marking a fundamental transition from programmed operations to true ambient intelligence.

Designed for deep domain integration, next-generation smart devices incorporate specialized sensor arrays and processing modules tailored to specific use cases. From precision image recognition to natural voice interaction, these purpose-built systems deliver customized intelligence while maintaining efficient operation. This trend reflects the maturation of embedded AI, marking a shift from merely receiving information and responding passively to actively perceiving and intelligently responding. Devices are increasingly capable of perceiving, analyzing, and adapting to their environments, enabling more natural and anticipatory interactions.

Classification of Smart Devices

Smart devices increasingly employ hybrid cloud-device architectures, using cloud collaboration for complex tasks while advancing localized processing through improved model compression and inference chips. This evolution delivers key consumer benefits: lower latency, stronger privacy, and more efficient interactions through edge computing.

Consumer-class smart devices[(1)] can be broadly categorized into three main categories based on their form and function. Smart home devices such as smart speakers and smart cameras, focus on voice interaction, security monitoring, and behavior perception within the household environment. Some of these devices perform initial processing locally before uploading complex data to the cloud for in-depth analysis. Wearable devices include AI-powered earphones, smart glasses, and smartwatches. They emphasize lightweight design and power-efficient inference, making them suitable for mobile environments where real-time perception and feedback are essential. Smart devices for children and learning devices are interactive devices with capabilities such as speech recognition and image processing that deliver intelligent experiences combining education and entertainment. These are tailored to specific needs such as early childhood learning and companionship.

Value Chain Analysis of Smart Devices

The value chain of smart devices comprises three key segments, including (1) upstream component suppliers, producing key elements that define device functionality and form factor, including core IC chips by semiconductor manufacturers, as well as sensors, display modules, batteries, and structural components, (2) midstream design and manufacturing specialists including Independent Design Houses, or IDH, for system integration and Original Design Manufacturers, or ODM, for production, and (3) downstream brand owners handling product definition, distribution, and consumer engagement, creating an efficient, specialized ecosystem from components to end-user experiences.

The value chain of smart devices is undergoing significant transformation as AI integration accelerates. Upstream component suppliers, including chip manufacturers and producers of other structural components are making breakthrough advancements in computing density and power efficiency, equipping devices with increasingly powerful on-device AI capabilities. These innovations

(1) Smart devices referenced here specifically refers to consumer-class products, purpose-built for targeted scenarios, excluding smartphones, personal computers, tablets, and automobiles. While these excluded categories also incorporate AI capabilities, they are primarily designed as multifunctional computing terminals, rather than smart devices such as AR glasses optimized for specialized intelligent functions.

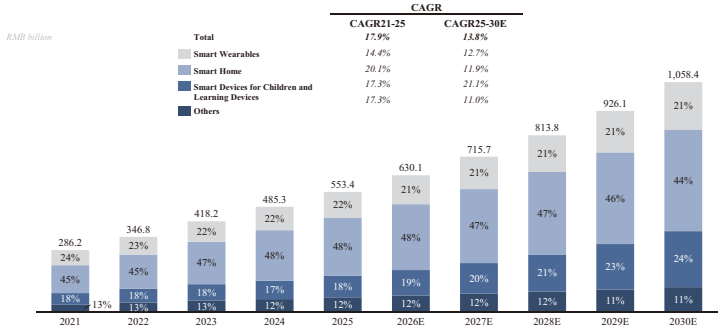
are enabling smarter, more autonomous functionality across consumer products. The midstream segment has become a strategic battleground, with IDHs and ODMs developing distinct value propositions. With growing focus on AI technology, the IDHs are emerging as a key competitive factor. IDHs now offer comprehensive solution services spanning hardware design, algorithm optimization, and system integration—delivering deeper technical expertise than traditional ODMs. Their ability to provide customized hardware-software co-development solutions enables strong product differentiation for specific industry applications, while ODMs focus on manufacturing excellence and rapid time-to-market. Brand owners are leveraging their market position to drive continuous innovation, using deep user understanding to guide product evolution. In an era of converging technologies and fragmented demand, the industry's progress increasingly depends on seamless cross-chain collaboration—combining upstream technical capabilities, midstream integration expertise, and downstream market intelligence to create winning smart device ecosystems.

Our main offerings consist of smart devices for consumer-class scenario, which serve both midstream and downstream segments.

Market Size of China's Smart Device Industry

China's smart device industry is experiencing rapid growth, particularly in consumer scenarios. The market size of China's consumer-class smart devices increased from RMB286.2 billion in 2021 to RMB553.4 billion in 2025, at a CAGR of 17.9%, and is expected to reach RMB1,058.4 billion by 2030 at a CAGR of 13.8% from 2025 to 2030.

Market Size of China's Consumer-Class Smart Device Industry, 2021–2030E



Source: Statista, CIC Report

Key Drivers and Trends for China's Smart Device Industry

Key growth drivers and trends of China's smart device industry include the following.

- *AI-driven transformation of smart devices.* The maturation of large AI models is fundamentally redefining smart device capabilities, extending advanced inference from the cloud to the edge through cloud-edge-device collaboration. AI-native products, as entirely new categories born from the advancement of AI technologies, are built around the capabilities of large models and offer intelligent services such as autonomous decision-making. In addition, traditional devices are being AI-empowered, enabling continuous functional enhancement. These devices are rapidly evolving from passive hardware into intelligent agents capable of proactively sensing user intent and delivering responsive services.

- *Regulatory framework and industry development.* China's expanding smart device market has seen progressive policy implementation, releasing an implementation plan for the digital transformation of the electronic information manufacturing industry (電子信息製造業數字化

轉型實施方案), which establishes robust requirements for data privacy, security, and interoperability. In September 2024, the MIIT issued a notice titled Guidelines for Advancing the Development of the Mobile IoT and the Vision of Everything Connected (關於推進移動物聯網「萬物智聯」發展的通知), further accelerating industry transformation by emphasizing the strategic value of integrated hardware-software solutions in driving demand for computing and storage services. These measures are elevating product quality, building consumer trust, and creating new value across the smart device ecosystem.

- *Domestic innovation and supply chain resilience.* Chinese smart device manufacturers are transitioning from import substitution to global technological leadership, supported by strong policy backing and domestic market demand. Breakthroughs in AI chips, proprietary algorithms, and system software have enabled Chinese firms to secure critical technological independence, significantly strengthening supply chain stability.

- *Advancing multimodal intelligence at the edge.* Smart devices are undergoing a fundamental transformation, evolving from basic sensors to sophisticated edge computing nodes. Through advancements in sensing technologies and AI algorithms, modern devices now integrate multiple input modalities, including visual, auditory, gestural, and emotional recognition, enabling more natural human-machine interactions. This shift from passive data collection to active environmental understanding and contextual response represents a quantum leap in device intelligence, making multimodal perception the new industry standard.

- *Rise of integrated solution offerings.* The industry is moving beyond standalone hardware to embrace holistic hardware-software-service business models. Forward-thinking IDH firms are developing proprietary algorithm platforms, management systems, and analytics backends that combine with their hardware expertise. These bundled solutions deliver significantly enhanced value through improved product functionality, stronger customer loyalty via recurring service revenue, and more sustainable business models that transcend traditional one-time hardware sales.

- *Ecosystem-centric development.* As device complexity and application scenarios multiply, seamless ecosystem integration has become critical. Leading enterprises now prioritize three key dimensions: tight hardware-software coordination, universal platform compatibility, and intelligent cross-device collaboration. This evolution from isolated products to interconnected systems enables unified management of device networks and coordinated responses across scenarios, dramatically improving both user experiences and operational efficiency while fostering a more vibrant smart device ecosystem.

COMPETITIVE LANDSCAPE

Only a selected group of industry leaders can provide integrated solutions across all the enterprise, consumer, and industrial applications. We stand among this elite tier as one of the top three industry leaders, in terms of revenue, uniquely positioned to deliver full-scenario Al inference chip-related products and services from the Chinese market.

The AI inference chip-related products and services industry broadly consists of two groups: independent vendors with full-stack capabilities and commercially available products, and providers offering internal business needs. We focus exclusively on the first group of independent players with AI inference chip-related products and services.

China's AI inference chip-related products and services industry is oligopolistic, with the two leading players collectively commanding approximately 83.6% market share in 2025. In 2025, we ranked No. 10 among all AI inference chip-related products and services providers from the Chinese market, in terms of revenue. The following table shows the ranking of top players in the industry.

Ranking of AI Inference Chip-Related Products and Services Providers from the Chinese market[(1)] by Revenue in 2025

Ranking	Company	Revenue from AI Inference Chip–Related Products and Services *(RMB billion)*	Market Share *(%)*	Main Technology Roadmap
1	Company A[(3)] ….	~200.0	~66.8%	GPU
2	Company B[(4)] ….	~51.1	~16.7%	NPU
3	Company C[(5)] ….	~5.5	~1.8%	NPU
4	Company D[(6)] ….	~3.0	~1.0%	XPU[(2)]
5	Company E[(7)] ….	~1.6	~0.5%	GPU
6	Company F[(8)] ….	~1.5	~0.5%	GPU
7	Company G[(9)] ….	~1.0	~0.3%	GPU
8	Company H[(10)] ….	~1.0	~0.3%	GPU
9	Company I[(11)] ….	~1.0	~0.3%	GPU
10	***Our Company ….***	***~0.7***	***~0.2%***	***NPU***

Source: Corporate Public Information and Annual Reports, CIC Report

(1) Excluding smart devices for consumer-class scenario.

(2) XPU is a form of GPU.

(3) Founded in 1993 and headquartered in the United States, company A is a fabless semiconductor company primarily engaged in the design and sale of GPUs. Company A is listed on Nasdaq.

(4) Company B is a private company founded in 1987 and headquartered in China. It is a global provider of information and communications technology (ICT) infrastructure and smart devices.

(5) Company C is founded in 2016 and headquartered in China. It is primarily engaged in the research and innovation of artificial intelligence chip products. It is listed on the Shanghai Stock Exchange.

(6) Company D is a private company founded in 2021 and headquartered in China. It specializes in the research, development and commercialization of AI chips and related hardware solutions.

(7) Company E is a provider of GPU chips and solutions headquartered in Shanghai and founded in 2020. It is a listed company on the Shanghai Stock Exchange.

(8) Company F is a full-function GPU provider, headquartered in Beijing and founded in 2020. It is a listed company on the Shanghai Stock Exchange.

(9) Company G focuses on the research and development of GPGPU chips and intelligent computing solutions, headquartered in Shanghai and founded in 2019. It is a listed company on the HKEX.

(10) Company H is a provider of GPGPU products and AI computing solutions headquartered in Shanghai and founded in 2015. It is a listed company on the HKEX.

(11) Company I is a provider of AI chips and computing solutions headquartered in Shanghai and founded in 2018. It is a private company.

In recent years, inference workloads have surged rapidly, becoming the main driving force behind AI applications. NPU, as a specialized and efficient technological solution, is better suited for inference tasks compared to GPU. We strategically positioned ourselves early in the NPU domain and have since established a strong market presence, ranking among the top three players in the industry. Players that entered the NPU domain early and have already secured leading positions in the market are likely to benefit more significantly from the accelerating penetration of NPUs, with greater potential for both market growth and share expansion.

Ranking of NPU-Powered AI Inference Chip-Related Products and Services Providers from the Chinese market[1] by Revenue in 2025

Ranking	Company	Revenue from AI Inference Chip-Related Products and Services	Market Share
		(RMB billion)	(%)
1	Company B	~51.1	~82.2%
2	Company C	~5.5	~8.9%
3	***Our Company***	***~0.7***	***~1.1%***

Source: Annual Reports, CIC Report

(1) Excluding smart devices for consumer-class scenario.

China's smart earphone module market remains highly concentrated, with the top five domestic manufacturers collectively holding 73.4% market share in 2025, while we lead the industry with a dominant 37.9% share by shipment volume, reflecting our strong position in this critical smart device category that enables voice interaction, real-time audio processing and seamless connectivity across devices.

Key Success Factors in China's Full-scenario Al Inference Chip-related Products and Services Industry

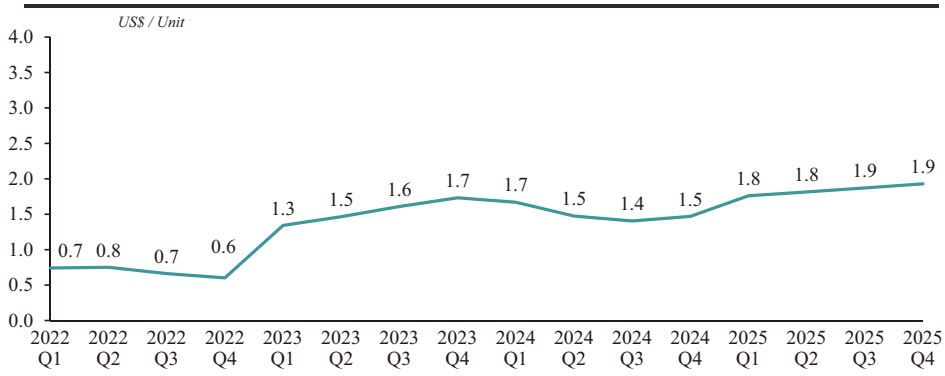
High technical threshold, long industry know-how accumulation cycle, talent advantage, customer advantage, fully localized supply chain, and full-stack solution and service capability are the key success factors of China's full-scenario AI inference chip-related products and services industry.

PRICE TREND OF APPLICATION-SPECIFIC INTEGRATED CIRCUITS

The primary raw materials for consumer-class products consist of chips and printed circuit boards (PCBs), in addition to common electronic components such as microphones, capacitors, and resistors. The chart below illustrates the global historical price trend of Application-Specific Integrated Circuits (ASICs) from Q1 2022 to Q4 2025. The price of ASICs in the Chinese market is also projected to remain generally stable.

The PCBs utilized in consumer-class products are typically customized to fit specific product architectures, with their cost influenced by board size, layer count, and processing complexity. Companies that source PCBs domestically tend to benefit from notable cost advantages due to localized procurement and more agile supply chains. As a result, these companies generally observe a stable to mildly declining unit cost trend for PCBs, barring major fluctuations in raw material prices or production capacity. Structural components are typically sourced from domestic suppliers and priced based on weight and material grade. These parts are often standardized and mass-produced, making them less sensitive to global commodity price fluctuations and contributing to consistently stable cost trends across the industry.

Global Average Selling Price of Application-Specific Integrated Circuits (by Quarter)



Source: WSTS, CIC report