The information and statistics set out in this section and other sections of this document were extracted from different official government publications, available sources from public market research and other sources from independent suppliers, and from the independent industry report prepared by China Insights Industry Consultancy Limited ("CIC"). We engaged CIC to prepare an independent industry report in connection with the [**REDACTED**] *(the "CIC Report"). The information from official government sources has not been independently verified by us, the Joint Sponsors,* [**REDACTED**]*, any of their respective directors, supervisors, and advisors, or any other parties involved in the* [**REDACTED**]*, and no representation is given as to its accuracy.*

OVERVIEW OF THE GLOBAL FOUNDATION MODEL INDUSTRY

Artificial intelligence (AI) is not only unlocking productivity, but also enriching creativity. Today, AI permeates various aspects of both people's professional and personal lives, from social media content recommendations, chatbots, and intelligent personal assistants, to autonomous driving systems, intelligent risk control models, and AI-assisted medical diagnostics. AI has emerged as a key enabler in the intelligent transformation of society and industries worldwide.

Foundation models in the past three years represent a significant technological paradigm shift compared to previous generations of AI — an inevitable trend fueled by societal developments. Traditional AI centered around small-scale models, which were custom-trained for application-specific scenarios. However, the goal is to enable intelligence that can perform the full range of human intellectual tasks. This requires AI to be more general-purpose, as user needs are becoming increasingly personalized and diverse. Foundation models are designed to address this challenge by offering scalability and generalization capabilities, and represent the most promising path towards achieving this goal.

Over the past three years, the field of foundation models has undergone rapid evolution. Benefitting from significant improvements in model scale and intelligence, the expansion of multi-modal capabilities, and the acceleration of commercialization, the industry has evolved at a remarkable pace.

SUSTAINED IMPROVEMENT IN MODEL INTELLIGENCE

Expansion of model scale and capabilities

Foundation models have scaled up dramatically in parameters in recent years, with significant performance improvements. OpenAI's GPT-3 launched in 2020 with 175 billion parameters, while GPT-4 launched in 2023 far outperformed it, scoring in the top 10% on a simulated bar exam versus GPT-3.5's bottom 10%, indicating near-human reasoning and comprehension capabilities.

Since then, more advanced models like GPT-4o, Gemini 2.5 Pro, and Claude 3.7 have pushed the boundaries of scale and intelligence further. A major breakthrough was the emergence of the Mixture-of-Experts (MoE) architecture, which uses expert sub-networks and a gating network to expand scale while keeping computational cost and latency low.

In 2025, leading foundation model companies accelerated model updates from once every 4 months or longer in 2024 to no more than 3 months, driving continuous improvements in intelligence. For example, in 2024, Anthropic launched the Claude 3 series in March, followed by the Claude 3.5 generation — including the Sonnet and Haiku lineups released in June and October respectively — which together constituted a major upgrade to the Claude 3 family. However, in 2025, Anthropic launched Claude 3.7 Sonnet in February, followed by Claude 4 just three months later in May, and Opus 4.1 in August, nearly 50% faster than the 2024 update pace.

Improving context windows and reasoning efficiency

New models are not only more intelligent, but also capable of memorizing and processing more content. GPT-3 supported a context window of approximately 2,048 tokens, GPT-3.5 increased to 4,096, and GPT-4 extended to 32,768. Claude's 100,000-token context window enabled interaction with ultra-long documents.

However, longer context windows raised inference costs, prompting architectural innovations and retrieval-augmented generation. Among the most notable innovations are improvements to the attention mechanism — for example, MiniMax Text-01, a prior version of our text model, marked the first large-scale application of the linear attention mechanism architecture.

Alignment with humans

RLHF (reinforcement learning from human feedback) allows foundation models to be more receptive to user prompts. It has now become a standard procedure that enhances instruction adherence and response quality. OpenAI used RLHF to evolve GPT-3.5 into ChatGPT, which is capable of delivering coherent, tailored, and practical answers.

This approach has also inspired alternatives like Anthropic's "Constitutional AI," which guides model behavior through predefined principles instead of human feedback. Human-aligned models show marked improvements in accuracy, tone control, and handling of inappropriate queries.

Emergence of CoT (chain-of-thought) and reasoning models

The CoT (chain-of-thought) prompting technique, introduced in 2022, improved performance on complex reasoning tasks, such as mathematical problem-solving, and common-sense reasoning, by generating intermediate steps.

A major shift in 2024 saw models trained to break down problems step-by-step during inference, allocating more compute to iterative reasoning, reflection, and output refinement. Reasoning is increasingly regarded as a computable process rather than merely an emergent ability due to large model size. Reasoning models built on test-time compute (such as OpenAI o1 and DeepSeek R1) are explicitly trained with additional compute during the reasoning process to conduct iterative or structured reasoning during test-time. This trade-off between increased reasoning cost/latency and higher reasoning quality may lead to a divergence in future model development: one class of models will be optimized for fast, factually accurate responses, while the other will focus on deeper, more resource-intensive reasoning, with test-time compute being a key variable.

Agentic tool use as a new paradigm

A new paradigm is emerging with AI agents — models that autonomously plan and use external tools to accomplish more complex tasks. In 2023, GPT-4 introduced plug-ins and function calling to access external tools like browsers or Python for code execution, overcoming the limitations of models that could previously only operate within the bounds of their training data. Gemini further advanced this by running code autonomously within a sandbox environment. In 2025, a number of leading foundation model companies have been focused on enhancing their models' agentic capabilities.

These capabilities turn models into intelligent agents, expanding their roles beyond passive response towards active task orchestration. Other tool use includes generating structured outputs for other systems to read and integrating with knowledge retrieval. Agentic AI has not only greatly expanded the scope of AI applications, but also represents a critical step forward.

Parallel development of closed- and open-source models

Both closed- and open-source models are advancing in parallel over the past few years. OpenAI released closed-source models like GPT-4, while Meta drove open-source adoption with the introduction of the open-sourced LLaMA 2, lowering the barrier of fine-tuned model development. The academic community also work together with the industry to accelerate research progress. Projects like Stanford's Alpaca and LMSYS's Vicuna democratized the exploration of instruction adherence models. The academic community also contributed to foundational model architecture innovations while helping advance model evaluation and safety.

The open-source momentum is pushing closed-source developers to iterate faster, while giving users more customizable model options. Chinese companies are also launching competitive open-source models, including Alibaba's Qwen3, DeepSeek's V3 and R1, and MiniMax's M1 and M2.

Acceleration of progress

The intelligence level of foundation models worldwide continues to advance. According to OpenAI's five-level roadmap, current models have now reached the threshold of Level 3. Looking ahead, the trajectory points clearly towards accelerated progress.

Levels	Name	Description
L1	Chatbots	AI with conversational language
L2	Reasoners	AI with human-level problem solving
L3	Agents	AI that can take actions
L4	Innovators	AI that can aid in invention
L5	Organizations	AI that can do the work of an organization

Source: OpenAI

CONTINUOUS EXPANSION OF MODALITIES

From single-modal to multi-modal

Foundation models have expanded into the multi-modal domain, aiming to integrate and align features from text, image, audio, and video into a shared semantic space, enabling integration across different modalities.

Visual understanding

In the early stages of multi-modal understanding, models like CLIP, ViLBERT, and VisualBERT primarily relied on dual-encoder architectures to align visual and textual inputs. More recently, the trend has been shifting towards more unified multi-modal capabilities. GPT-4V, for example, extends the GPT-4 framework to support image inputs, allowing users to ask the model to analyze visual content, describe image details, interpret humor in memes and information in medical images. Built on a decoder-only architecture, Gemini supports image, video, and audio modalities, with Gemini Ultra setting new benchmarks in multi-modal reasoning tasks.

Audio generation

The integration of text and audio allows AI to interpret and generate audio itself.

OpenAI's Whisper, launched in 2022 with 1.6 billion parameters, transcribes and translates audio in 97 languages, achieving near-human accuracy in English transcription. It enables developers to convert audio into text for further processing.

Audio synthesis has also advanced rapidly. In 2023, service providers such as ElevenLabs and MiniMax enabled models to speak with human-like voices. That same year, OpenAI added audio-based conversation to ChatGPT, allowing real-time spoken input and synthesized output, expanding the application of audio model in intelligent assistants and customer service.

The integration of text and audio has led to new products like voice-driven AI agents and smart devices. Future models will better understand emotion and intent in speech, producing more natural responses and improving human-machine interaction.

Visual generation

By 2022, text-to-image models such as DALL-E, Imagen, Stable Diffusion, and Midjourney began producing outputs comparable to real photos and artwork. These models are typically based on diffusion models, combining language models with generative models based on large image-text datasets and Transformer-based text encoding. DALL-E 3 supports natural language interactions and can edit in-image text. Stable Diffusion is noted for photo realism, customization, and open-source engagement. Midjourney excels in artistic styling and usability.

Since 2023, video generation has emerged as a new multi-modal space. OpenAI's Sora, a DiT-powered video model, can generate new video content from inputs in the forms of text, image, or even video. Other offerings, such as Hailuo AI and Google Veo 3, have also gained global traction. These tools have democratized creative content generation and improved workflow efficiency in the creative industry.

The academic community has begun exploring unified models capable of both multi-modal understanding and generation. These models are designed to handle diverse input modalities and generate outputs across one or more of those modalities within a single, cohesive architecture. Such unified systems need to combine the advantages of autoregressive models in reasoning and text generation, with the strength of diffusion models in high-fidelity image generation. This pursuit of integration mirrors the deeper nature of human intelligence — human understanding and expression, as well as inputs and outputs across different modalities, are deeply intertwined and inseparable, rather than being modular and independent of each other.

RISING ADOPTION OF FOUNDATION MODEL APPLICATIONS UNLOCKING COMMERCIAL VALUE

Unprecedented growth of foundation model applications

Over the past three years, the new generation of AI has experienced hyper growth, at a rate surpassing all previous technological waves in human history, such as the internet and the industrial revolution.

ChatGPT became the fastest growing product in history to reach 800 million users, taking only 17 months, while achieving a global reach with more than 90% of users from outside North America. Commercially, the new generation of AI-native products reached levels of revenue in a single year that took the SaaS industry a decade to accomplish. Powered by the infrastructure and momentum of previous technological waves, AI technology is spreading across every corner of the internet at lightning speed.

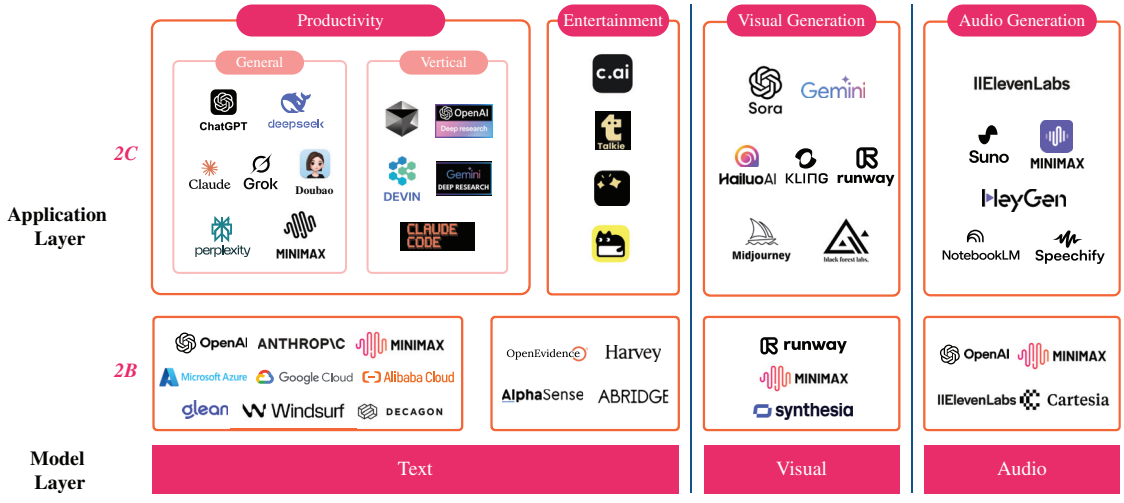
Humanity is now at a pivotal inflection point of an exponential technological growth trajectory. In the moment, progress may seem linear to people, though looking back, exponential leaps often unfold in a very short span of time.

Massive TAM of foundation model applications driven by generalization

Currently, major applications of foundation models include productivity, entertainment, visual generation, audio generation and general 2B services. Across these segments, the generalization capabilities of foundation models are enabling both large-scale deployment and personalized applications through a single, highly scalable architecture—serving a broad spectrum of user needs, from large enterprises to individual creators, and delivering positive model ROI.

Products that can achieve sustained organic growth over time are those driven by continuous improvements in underlying model intelligence. Players that intend to maintain long-term market leadership must develop proprietary foundation models that can be optimized from end-to-end and consistently maintain top-tier performance. A breakthrough in foundation model intelligence can rapidly propel them to prominence, as users naturally gravitate towards technologies that offer a more positive experience.

Market landscape of foundation model applications



Source: CIC

Productivity

The productivity segment represents a massive opportunity with broad downstream use cases. Users often engage with multiple models simultaneously. Top use cases include information search, writing, coding, education, office administration, academic research, and business analysis, which cover all aspects of work and daily life.

The productivity segment has undergone a generational shift from chatbots in 2023 to agents in 2025. Leading chatbot players include ChatGPT and DeepSeek R1, with OpenAI Deep Research and MiniMax Agent driving the next generation of AI agents. Unlike chatbots that simply respond to prompts, agents can complete long-horizon tasks due to advances in multi-step reasoning and use of external tools, enabling them to learn and improve through interactions with their environment. An emerging application of agents is virtual co-workers that can be integrated into enterprise workflows. Companies such as OpenAI and Anthropic are training foundation models in reinforcement learning environments to operate professional business software. The ultimate goal is for these agents to independently handle complex tasks and deliver tangible business value. Long-term leaders in this field must possess end-to-end capabilities, the ability of models to enhance their capabilities via end-to-end reinforcement learning using proprietary models and rewards from application-specific environments.

Since the second half of 2024, AI coding applications experienced exponential growth, fueled by the breakthrough of Claude 3.5 Sonnet. Its capabilities in code design, debugging, and optimization have powered over 30 million developers worldwide. Notable products include Claude Code, Anthropic's agentic coding tool, Cursor, a code editor for professional developers, and Windsurf, an enterprise-level secure coding platform. Beyond the professional coding market, 2025 has seen a surge of "vibe coding" tools designed for everyday users with no programming background. Platforms such as Lovable and Bolt.new enable anyone to create applications simply through natural language input. The overall trend is shifting from simple code completion towards more advanced coding agent capabilities, with a long-term potential for enabling personalized software generation from a single chat interface. This would not only lower barriers for professional product development, but also unlock a new market for users with little experience.

Entertainment

Entertainment is the second-largest segment following productivity, with tens of millions of young users worldwide creating and interacting with personalized AI agents. The use cases are highly diverse, spanning role-play, companionship, and a wide range of everyday Q&A interactions.

Competition in the entertainment segment is in the process of stabilizing. Leading products include Character AI and Talkie/Xingye, which can enhance user experience with model optimization and inspire users' creativity with rich multi-modal creative tools. This combination drives high engagement, user stickiness, and highly interactive experiences.

The new generation of AI-native users are naturally inclined to interact with AI companions. As societal productivity continues to rise and material needs are increasingly met, entertainment AI products will tap into users' emotional and psychological needs and unlock long-term market potential through personalized, emotionally resonant experiences.

Visual generation

Image generation has emerged as the first AI domain to achieve commercialization. In 2022, models like Midjourney impressed the world with their visually striking outputs, sparking exponential growth in social media engagement as image quality progressively met commercial standards. The primary user base consists of professional creators and enthusiasts in graphic design, film/TV production, advertising and e-commerce. These tools not only inspire creativity but also significantly enhance design workflow efficiency.

Leading applications in this space, including GPT-4o, Midjourney and Flux, continue to make breakthroughs in image quality, editing customizability, and diversity of styles. In 2025, Google's Nano Banana advanced image generation to a new level, enabling precise natural language editing and powering commercial-grade uses from marketing visuals to game design. These advancements have unlocked greater end-user application scenarios, such as professional-standard product design and commercial marketing materials. AI-generated images have already achieved widespread popularity, marking their evolution from being just creative tools to becoming mainstream content.

Video generation has emerged as a rapidly growing segment in 2024, with a clear product-market fit. Demand comes from a wide range of industries, including film and television, short videos, mini-dramas, advertising, and e-commerce, leading to a massive market opportunity. In these industries, conventional video production often requires an entire team, whereas foundation models open up new market opportunities for individual professional creators to act as "one-person studios" to produce high-value content as well as enhancing their productivity.

Leading players in this segment include Sora, Veo, Hailuo AI, and Kling, among others. Their core competitiveness lies in maintaining our model R&D capabilities and cost efficiencies, coupled with fast-iterating creative workflow features and a vibrant creator ecosystem.

AI-generated videos are beginning to go viral increasingly frequently on social media, signaling that model performance is beginning to break through the boundaries of consumer-level content. Sora 2, launched in October 2025, sparked viral sharing on social media, signaling a major shift in the content industry with opportunities on the scale of the next short-video boom. In the future, relevant products may evolve into a "real-time personalized video generation engine", lowering barriers for anyone to create and consume personalized content.

Audio generation

Audio is the universal interface of interaction in the AI era, with a broad downstream application market. For enterprises, AI voice agents overcome the limit of human capacity in sales and customer service, including recruitment, finance, healthcare; for content creators, it enables lifelike and emotionally expressive audio generation for audiobooks, education, dubbing and gaming, and others.

Leading players include OpenAI, MiniMax, and ElevenLabs. Their core competitiveness lies in delivering hyper-realistic audio model quality while maintaining low cost and low latency.

Numerous agentic AI applications, and smart devices are empowered by audio in the AI era. OpenAI's GPT-4o introduced real-time audio interaction in May 2024, setting a new standard for chatbots; Google's NotebookLM saw viral success in September 2024 with its podcast generation feature. As human—AI interactions grow exponentially, the audio submarket holds vast untapped potential.

General 2B services

To accelerate AI adoption in various fields, foundation model companies such as OpenAI and Anthropic typically offer model capabilities to developers and enterprise clients via APIs with an open-platform strategy. Cloud service providers such as Microsoft, Amazon, Google, and Alibaba also provide models, toolkits and professional services through APIs, industry-tailored solutions, and on-premise deployment.

The core competitiveness in this segment includes model performance, cost-efficiency, and stability during high concurrencies, which are the top concerns for developers and enterprise customers. Secondary considerations include security, compliance, and customer support. A multiple-model strategy is now common, with enterprises often using three or more models and routing different models to specific tasks based on use-case requirements.

Enterprise demand is surging across industries. As agentic models become increasingly capable of delivering satisfying outcomes and inference costs continue to drop rapidly, foundation models are set to become a new productivity norm, continuously unlocking value across sectors.

SCALE OF THE GLOBAL FOUNDATION MODEL MARKET

Market size

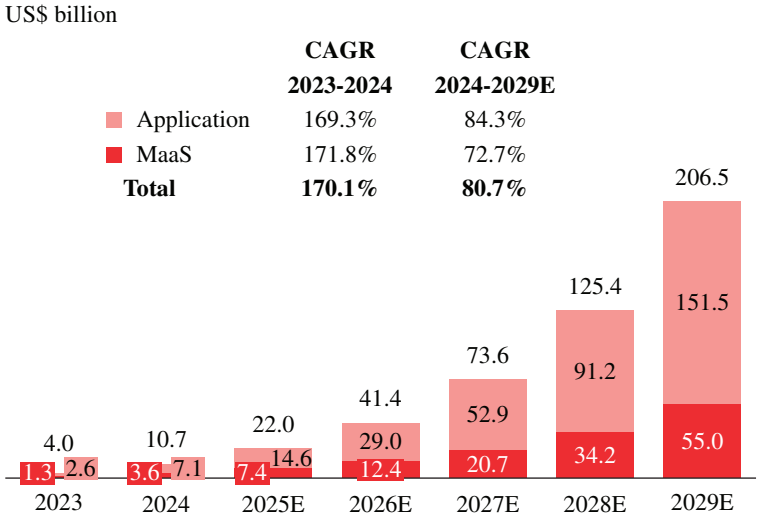
The global foundation model market comprises revenue generated by model-based and deployment-based approaches. Model-based revenue is primarily generated from (i) a wide range of end-user applications such as AI chatbots, social and entertainment AI products, video generation and audio generation products, that are offered to both consumers and enterprises

mainly via subscriptions, and (ii) MaaS (model-as-a-service), which refers to the provision of foundation model capabilities via cloud-based APIs and licensing, enabling developers and enterprises to access and integrate model functions into their own products or systems on a usage basis. Deployment-based revenue is generated from the deployment of customized solutions on premise.

Foundation model technology remains in a stage of rapid development. Compared to a deployment-based approach, the model-based approach allows users to benefit from continuous model improvements without incurring version migration costs. Users can also dynamically scale their model usage based on actual demand, reducing upfront investments and ongoing maintenance expenses related to hardware and infrastructure. Moreover, this approach supports automatic resource scaling to meet users' evolving needs.

According to CIC, the global model-based foundation model market is still in the early stages of commercialization. As technologies continue to mature and the willingness of users to pay steadily increases, the global model-based foundation model market is expected to grow rapidly from US$10.7 billion in 2024 to US$206.5 billion by 2029, representing a CAGR of 80.7%. Driven by continued advancement and maturity of foundation model technologies, the market size of foundation model application is projected to expand from US$7.1 billion in 2024 to US$151.5 billion in 2029, at a CAGR of 84.3%, and the market size of foundation model MaaS is expected to grow from US$3.6 billion in 2024 to US$55.0 billion in 2029, representing a CAGR of 72.7%.

The global foundation model market size, in terms of model-based revenue, 2023-2029E



Source: CIC

Note: Model-based revenues primarily include income generated from foundation model application subscriptions, and foundation model API calls and licensing.

Market drivers

Technological leaps

The foundation model market is characterized by disruptive technological breakthroughs, with the improvements in each new generation of foundation models expanding the scope of potential applications.

GPT-3 enabled entertainment chat to first achieve product-market fit; GPT-3.5 brought chatbot applications to a highly usable level, while GPT-4 tapped into professional domains such as finance and law. Sora drove video generation to meet the commercial requirements for quality. The multi-modal GPT-4o facilitated a new surge in new user adoption for ChatGPT. Claude 3.5 Sonnet's enhanced coding capabilities contributed to a product-market fit for developer tools like Claude Code, Cursor and Windsurf. OpenAI's o1, with improved reasoning, and Claude 3.7 Sonnet, with stronger tool use capabilities, are fueling the rise of AI agents.

Technology serves as the most powerful underlying driver to the wave of foundation models, with foundation model companies positioned at the forefront of this transformative growth. Each new generation of models gives rise to new use cases that evolve from fragmented experimentation into mainstream applications. True breakthroughs typically occur between generations, with each leap opening a new capability curve and enabling entirely new categories of products and services. For example, MiniMax-M2, MiniMax's latest text model, incorporated "interleaved thinking", a novel, non-consensus framework that enables more robust and reliable agentic reasoning and has proven highly effective. Within the first week of its launch, it became a top three foundation model worldwide by daily token usage on OpenRouter (one of the most widely used platforms globally that lets developers easily access multiple foundation models through one unified API), as well as the first China-based model to surpass 50 billion daily token usage on OpenRouter.

For foundation model companies, these inflection points directly translate into the expansion of demand for new products and solutions. By exploring the direction of model evolution, they would be able to align product roadmaps with emerging market needs, accelerate customer adoption, and capture growth at the earliest stages of each technology cycle.

Scaling Law

The fundamental driver behind market growth of foundation models lies in the fact that the foundation model technology is able to keep scaling up.

The pre-training scaling law is well-known — model performance improves in proportion to the increases in model scale, data size and computing power. This law still remains valid across models from GPT-3.5, GPT-4, to the recently released Claude 4, whose parameter numbers expanded from hundreds of billions to trillions. Furthermore, this principle applies not only to text models but also to other modalities such as video and audio, as recent breakthroughs in video and audio generation technologies have similarly benefited from the scaling up of both model scale and training data size.

Moreover, beginning with OpenAI's o1, the industry witnessed a new scaling law focused on test-time compute. The model's reasoning capability enhances as the computational load in inference extends — the longer the time that the model spends on thinking, the better the performance. Long thinking can require 100 times more compute than a single inference session, allowing the model to solve incredibly complex tasks that can barely handled by conventional models. This principle has been consistently validated in subsequent models such as OpenAI's o3, and as of November 2025, all the top ten best-performing models in terms of intelligence index on Artificial Analysis are reasoning models, including MiniMax-M2.

Looking ahead, the scaling up of foundation models is expected to continue, with the scaling of both pre-training and inference reinforcing each other. This dynamic underpins the new "Moore's Law," which benefits the entire industry through collective scaling progress rather than isolated innovation. It allows foundation model companies to deploy increasingly sophisticated models with higher throughput and lower latency, and capture growth associated with the scaling trajectory of the industry.

Cost reduction

Declining model costs represent a more predictable market driver than the improvements in model capabilities, with both factors expected to unlock an increasing number of use cases that are crossing the ROI threshold and achieving product-market fit.

At the time of GPT-4's release, many vertical applications, such as content moderation, already met performance requirements but remained commercially unviable due to high costs and negative ROI. The inference cost of foundation models has been decreasing steadily, with per-token cost of GPT-4 dropping by over 99% since its release, enabling broader adoption across high-volume, back-end industry scenarios. This decline is consistently observed in the industry and has been driven by a combination of architecture innovations, inference efficiency improvements, engineering optimizations, and reductions in the cost of compute. These factors are expected to continually lower costs at a predictable rate.

Falling inference costs expand the range of economically viable applications, lowering adoption barriers and unlocking new market opportunities. This trend drives higher inference volumes and broader deployment of foundation model products and solutions, enabling customers to scale usage profitably and accelerating overall market growth.

Trends of foundation model applications

The commercialization of foundation model applications is still in its early stages, with the proximity to agents being the inflection point that is positioned to unleash significant commercial values.

Agent applications: agents are capable of operating at a professional level, acting autonomously, and delivering end-to-end results, ultimately driving GDP of trillions of dollars' worth

Achieving professional-level performance involves the execution of specialized tasks within expert domains. Acting autonomously allows systems to operate independently of human time and attention. Delivering end-to-end results signifies the ability to generate economic value. These characteristics mark the inflection point for LLMs transitioning from offering tools to delivering results. Consequently, the addressable market for LLMs will expand beyond the boundaries of enterprise software budget and into the broader market space for labor services.

In addition, agents are continuously enhancing their capabilities to complete tasks. Agents also follow its own scaling law —with the duration of tasks that agents can autonomously handle doubling approximately every seven months. Today, AI can autonomously complete tasks that typically take humans one hour to complete. AI is expected to autonomously handle 2-3-hour tasks by 2026, and one-month tasks within five years. This paves the way for a future of "infinite experts" — AI software engineers, financial analysts, and research scientists contributing to greater productivity and agent economy, 24/7 without downtime.

In light of the agent scaling law, future agents will not only execute tasks, but also act as AI researchers that are able to accelerate the research of new proprietary algorithms and develop new agents that surpass their own capabilities. This exponential, self-reinforcing evolution is precisely what sets this generation of AI technology apart at its core. A compounding effect will emerge between the scaling of algorithm and application, each reinforcing and accelerating the other. This feedback loop could make AI the fastest-moving technological revolution in human history.

Entertainment and generative applications: rapid growth across multiple verticals

The new generation of AI-native users seeks immersive, co-creation experiences, driving the evolution of entertainment products toward personalized AI companionship. As model intelligence and memory capabilities continue to advance, we foresee a future "Her"-style moment where everyone has an AI companion that truly knows them and proactively assists in all aspects of their lives. These AI companions will possess both intellectual and emotional intelligence, forging deep emotional bonds with users through expressive and empathetic interactions. Personalized AI emerges as a trend where AI companions can learn user's personalized preferences, habits, and communication style from daily interactions, assisting them across all devices.

Advances in video generation are altering the limits of creativity, as AI-generated videos have the potential to become viral on social media. This indicates a shift in content production from professional tools to widely accessible creative engines, significantly changing the video content supply landscape. Meanwhile, audio generation and interaction capabilities are

gradually becoming standard across AI applications, advancing from basic command-response functions toward more emotionally expressive communication. This evolution is enabling more natural interactions between human and machines, positioning the voice interface as a central hub for multi-modal interaction.

Multi-modal applications: unified multi-modalities unlocks new market potential

In March 2025, GPT-4o updated image generation capabilities, significantly improving image quality and triggering a sharp uptick in new ChatGPT subscriptions, demonstrating the commercial potential of multi-modal integration. Unlike previous approaches that relied on separate models like DALL-E for image generation, GPT-4o is built on a native multi-modal architecture that generates image directly from text prompts. This has brought a new level of controllable image generation and editing.

The ability to accurately generate text within images and to edit images with fine-grained control has opened up commercial use cases such as generating educational visuals, product posters with stylized typography, and scientific illustrations. Looking ahead, deeper integration of text, audio, and visual modalities will make it possible to create fully editable videos, generate synchronized audio and text along with the video content, and more, unlocking market opportunities for the next short-video revolution.

COMPETITIVE LANDSCAPE OF THE GLOBAL FOUNDATION MODEL MARKET

Competitive ranking

Foundation model companies are broadly categorized into two types: foundation model technology companies and foundation model application companies. The former refers to companies capable of developing proprietary foundation model technology, while the latter refers to those building industry- or scenario-specific applications and solutions on top of existing foundation models, without engaging in foundation model development or maintenance themselves.

Currently, the foundation model industry is still in a phase driven by the advancements of underlying technology, where major iterations of foundation models can significantly expand the boundaries of model capabilities. As a result, foundation model technology companies who are more focused on advancing the underlying technology are the major promoters of innovation and play a leading role in shaping the future of the industry.

As end-user experience is largely dependent on the performance of the foundation model, a large number of leading products in the market today are developed by foundation model technology companies with end-to-end model and application development capabilities.

According to CIC, MiniMax is the tenth largest foundation model technology company globally in terms of model-based revenues in 2024, with a market share of 0.3%, as illustrated in the table below. The global foundation model market is expected to reach US$22.0 billion

in 2025, and MiniMax is expected to capture a market share of approximately 0.3%. Given that most of the peers are large publicly listed companies with vast resources, ranking among the global top ten and competing effectively with these industry giants is a remarkable achievement for a startup with relatively limited resources.

Ranking of global foundation model technology companies, in terms of model-based revenues in 2024

Rank	Company	Market share, %
1	Company A	30.1%
2	Company B	16.9%
3	Company C	8.2%
4	Company D	4.7%
5	Company E	2.8%
6	Company F	1.8%
7	Company G	0.7%
8	Company H	0.5%
9	Company I	0.3%
10	**MiniMax**	**0.3%**
11	Company J	0.3%
12	Company K	0.3%
13	Company L	0.3%
14	Company M	0.2%
15	Company N	0.2%

Source: CIC

Note:

(1) Model-based revenues primarily include income generated from foundation model application subscriptions, and foundation model API calls and licensing.

(2) Company A is a foundation model company founded in the United States in 2015. It mainly provides AI-native products such as chatbot and video generation application. It is an unlisted company.

(3) Company B is a technology company founded in the United States in 1998. It mainly provides internet-related products and services, including search engines, cloud computing, digital advertising, and AI products and services. It is a listed company on the NASDAQ Stock Exchange.

(4) Company C is a technology company founded in the United States in 1975. It mainly provides office software, cloud services, and AI products and services. It is a listed company on the NASDAQ Stock Exchange.

(5) Company D is a foundation model company founded in the United States in 2021. It mainly provides large language model products. It is an unlisted company.

(6) Company E is a foundation model company founded in the United States in 2021. It mainly provides AI image generation application. It is an unlisted company.

(7) Company F is a technology company founded in the United States in 1994. It mainly provides an e-commerce platform, cloud computing services, digital streaming, and AI products and services. It is a listed company on NASDAQ Stock Exchange.

(8) Company G is a technology company founded in the United States in 2004. It mainly provides social networking platforms, and open-source foundation models. It is a listed company on NASDAQ Stock Exchange.

(9) Company H is a technology company founded in China in 1999. It mainly provides e-commerce platforms, cloud computing services, digital payment services, and AI products and services. It is a dual-listed company on the Stock Exchange and the New York Stock Exchange.

(10) Company I is a social media company founded in the United States in 2006. It mainly provides a global social networking platform and a large language model application. It is an unlisted company.

(11) Company J is an AI company founded in the United States in 2022. It mainly provides AI-powered voice synthesis and dubbing services, including multilingual speech generation and voice cloning. It is an unlisted company.

(12) Company K is a technology company founded in China in 2000. It mainly provides search engine, cloud services, and AI products and services. It is a dual-listed company on the Stock Exchange and the NASDAQ Stock Exchange.

(13) Company L is an AI company founded in the United States in 2018. It mainly provides an AI-powered video and image generation tools. It is an unlisted company.

(14) Company M is an AI company founded in the United Kingdom in 2017. It mainly provides AI-powered video creation tools. It is an unlisted company.

(15) Company N is a voice technology and AI company founded in China in 1999. It mainly provides voice recognition software and other voice-based AI products. It is a listed company on the Shenzhen Stock Exchange.

The following table presents a comparison of product offerings by leading global foundation model technology companies.

Product offerings by leading global foundation model technology companies

Company	Main product types	Main monetisation method
Company A	Productivity, visual generation	Subscriptions, API calls
Company B	Productivity, general 2B services	Subscriptions, API calls
Company C	Productivity, general 2B services	Subscriptions, API calls
Company D	Productivity	Subscriptions
Company E	Visual generation	Subscriptions
Company F	General 2B services	API calls
Company G	General 2B services	API calls
Company H	General 2B services	Subscriptions, API calls
Company I	Productivity, general 2B services	Subscriptions, API calls

Company	Main product types	Main monetisation method
MiniMax	Entertainment, visual generation, productivity	Subscriptions, online marketing services, in-app purchase, API calls
Company J	Audio generation	Subscriptions, API calls
Company K	Productivity, general 2B services	Subscriptions, API calls
Company L	Visual generation	Subscriptions, API calls
Company M	Visual generation	Subscriptions, API calls
Company N	Productivity, general 2B services	Subscriptions, API calls

The table below summarizes the key underlying technologies used by leading foundation model technology companies across text, image, video, and audio modalities, as a high-level representation of each company's technical orientation.

Underlying technologies used by leading global foundation model technology companies

	Model Modalities			
Company	Text	Image	Video	Audio
Company A. . .	RLHF, SFT, RAG, CoT, MoE	Diffusion model, Cross-Attention, MoE, ViT	Spatiotemporal Attention, Video Diffusion Model, Multimodal Fusion, MoE	TTS, Codec-based Model, Diffusion-based Vocoder, MoE
Company B. . .	RLHF, SFT, RAG, CoT, MoE	Diffusion model, Cross-Attention, MoE, ViT	Spatiotemporal Attention, Video Diffusion Model, Multimodal Fusion, MoE	TTS, Codec-based Model, Diffusion-based Vocoder, MoE
Company C. . .	RLHF, SFT, RAG, CoT, MoE	Diffusion model, Cross-Attention, ViT	Spatiotemporal Attention, Video Diffusion Model, Multimodal Fusion	TTS, Codec-based Model, Diffusion-based Vocoder, MoE
Company D. . .	RLHF, SFT, CoT, Interleaved Thinking	N.A.	N.A.	N.A.
Company E. . .	N.A.	Diffusion model, Cross-Attention	N.A.	N.A.
Company F. . .	RLHF, SFT, RAG, CoT, MoE	Diffusion model, Cross-Attention, MoE, ViT	Spatiotemporal Attention, Video Diffusion Model, Multimodal Fusion, MoE	TTS, Codec-based Model, Diffusion-based Vocoder, MoE

	Model Modalities			
Company	Text	Image	Video	Audio
Company G. . .	RLHF, SFT, CoT, MoE	Diffusion Model, Cross-Attention, MoE, ViT	Spatiotemporal Attention, Video Diffusion Model, Multimodal Fusion, MoE	TTS, Codec-based Model, Diffusion-based Vocoder, MoE
Company H. . .	RLHF, SFT, RAG, CoT, MoE, Linear Attention	Diffusion Model, Cross-Attention, MoE, ViT	Spatiotemporal Attention, Video Diffusion Model, Multimodal Fusion, MoE	TTS, Codec-based Model, Diffusion-based Vocoder, MoE
Company I . . .	RLHF, SFT, RAG, CoT	Cross-Attention, ViT	N.A.	N.A.
MiniMax	RLHF, SFT, RAG, CoT, MoE, Linear Attention, Interleaved Thinking	Diffusion Model, Cross-Attention, MoE, ViT	Spatiotemporal Attention, Video Diffusion Model, Multimodal Fusion, MoE	TTS, Codec-based Model, Diffusion-based Vocoder, MoE
Company J . . .	N.A.	N.A.	N.A.	TTS, Codec-based Model, Diffusion-based Vocoder
Company K. . .	RLHF, SFT, RAG, CoT, MoE	Diffusion Model, Cross-Attention, MoE, ViT	Spatiotemporal Attention, Video Diffusion Model, Multimodal Fusion, MoE	TTS, Codec-based Model, Diffusion-based Vocoder, MoE
Company L. . .	N.A.	N.A.	Spatiotemporal Attention, Video Diffusion Model, Multimodal Fusion	TTS, Codec-based Model, Diffusion-based Vocoder
Company M . .	N.A.	N.A.	Spatiotemporal Attention, Video Diffusion Model, Multimodal Fusion	TTS, Codec-based Model, Diffusion-based Vocoder
Company N. . .	RLHF, SFT, RAG, CoT	N.A.	N.A.	TTS, Codec-based Model, Diffusion-based Vocoder, MoE

Notes:

1. RLHF (Reinforcement Learning from Human Feedback) refers to a training method where the model learns to produce responses that align with human preferences, by being rewarded for outputs that human rate as better.

2. SFT (Supervised Fine-Tuning) refers to a process of improving a model by training it on high-quality examples with known correct answers, so it learns to imitate desired behavior.

3. RAG (Retrieval-Augmented Generation) refers to a technique where the model retrieves relevant information from external databases or documents to generate more accurate and factual answers.

4. CoT (Chain-of-Thought reasoning) refers to a reasoning approach where the model generates intermediate thinking steps before producing the final answer, improving logical accuracy and problem-solving.

5. MoE (Mixture-of-Experts) refers to a model architecture that contains multiple specialized "expert" networks, where only the most relevant ones are activated for each input, improving efficiency and scalability.

6. Linear Attention refers to an optimized form of attention that reduces memory and computation costs, enabling the model to handle much longer input sequences efficiently.

7. Interleaved Thinking refers to a reasoning approach where a model alternates between multiple lines of thought or tasks, allowing it to process complex problems more efficiently and generate more coherent, context-aware outputs.

8. Diffusion Model refers to a generative approach that starts from random noise and progressively refines it into a clear image, similar to developing a photograph.

9. Cross-Attention refers to a mechanism that enables a model to connect and align information from different sources, such as linking text prompts to visual features.

10. ViT (Vision Transformer) refers to a Transformer architecture designed for image understanding, which divides an image into small patches and processes them to capture global visual patterns.

11. Spatiotemporal Attention refers to an attention mechanism that jointly analyzes spatial information (objects in each frame) and temporal information (how things move across frames) to understand videos.

12. Video Diffusion Model refers to a generative framework that extends diffusion models to videos, creating smooth and coherent motion by refining noisy video frames step by step.

13. Multimodal Fusion refers to the process of combining multiple types of data — such as text, image, audio, and video — so that the model can understand or generate content across modalities.

14. TTS (Text-to-Speech) refers to the process of converting written text into spoken voice, allowing machines to "speak" naturally.

15. Codec-based Model refers to an audio generation approach that compresses sound into compact digital codes (tokens) and reconstructs it with high fidelity, similar to how MP3 or EnCodec works.

16. Diffusion-based Vocoder refers to a model that reconstructs realistic audio waveforms from encoded representations through a gradual noise-removal process, improving speech and music quality.

Foundation model technology companies can be further classified into two categories: pureplay companies whose core business is entirely focused on foundation models, and non-pureplay companies that have entered the foundation model space in addition to their existing businesses, such as major internet platforms and cloud service providers.

Pureplay companies concentrate their core resources, accumulated technological know-how and business models around foundation models. This high degree of focus and resource investment enables them to drive rapid innovation and positions them as key forces in advancing the foundation model industry. In contrast, non-pureplay companies may benefit from stronger access to capital and computing power. They are also able to integrate foundation model technologies into a wider range of products and services from other business units or departments across the organization, enabling potentially faster and easier validation and commercialization of new technologies.

According to CIC, MiniMax is the fourth largest pureplay foundation model technology company globally in terms of model-based revenues in 2024, as is illustrated in the tables below.

Ranking of global pureplay foundation model technology companies, in terms of model-based revenues in 2024

Rank	Company	Market share, %
1	Company A	30.1%
2	Company D	4.7%
3	Company E	2.8%
4	**MiniMax**	**0.3%**
5	Company J	0.3%

Source: CIC

Note:

(1) Model-based revenues primarily include income generated from foundation model application subscriptions, and foundation model API calls and licensing.

Model benchmarks

Upon release, MiniMax's foundation models have achieved leading performance across text, video and speech modalities, ranking at the top across the Artificial Analysis benchmarks and achieved the No.1 ranking among all open-source models, a suite of authoritative, independent AI benchmarks that are widely acknowledged in the foundation model industry, providing assessments from the perspective of large model users, as illustrated in the charts below.

Artificial Analysis Intelligence Index (evaluation of text models)

Rank	Company	Model	Index
1	OpenAI	GPT-5 Codex (high)	68
1	OpenAI	GPT-5 (high)	68
3	X	Grok 4	65
4	Anthropic	Claude 4.5 Sonnet	63
5	**MiniMax**	**MiniMax-M2**	**61**
5	OpenAI	gpt-oss-120B (high)	61
7	X	Grok 4 Fast	60
7	Google	Gemini 2.5 Pro	60
9	Anthropic	Claude 4.1 Opus	59
10	Alibaba	Qwen3 235B A22B 2507	57

Source: Artificial Analysis

Note: As of November 7, 2025, shortly after the release of MiniMax-M2. Artificial Analysis is an independent AI benchmarking & analysis company. It provides independent benchmarks & analysis to support developers, researchers, businesses, and other users of AI. The Artificial Analysis Intelligence Index is a weighted average metric across the constituent evaluations, balancing general knowledge (equally weighted between MMLU-Pro, HLE, and GPQA Diamond), mathematical reasoning (equally weighted between MATH-500 and AIME 2024), and coding ability combination (equally weighted between SciCode and LiveCodeBench).

Artificial Analysis Video Arena Leaderboard (evaluation of video models)

Rank	Company	Model	Arena ELO
1	ByteDance	Seedance 1.0	1,355
2	**MiniMax**	**Hailuo-02**	**1,331**
3	Google	Veo 3 Preview (No Audio)	1,244
4	Kuaishou	Kling 2.0	1,195
5	Kuaishou	Kling 1.6 (Pro)	1,144
6	Runway	Runway Gen 4	1,120
7	Google	Veo 2	1,118
8	Lightricks	LTV Video v0.9.7 (13B)	1,064
9	MiniMax	I2V-01-Director	1,047
10	Runway	Runway Gen 3 Alpha Turbo	1,005

Source: Artificial Analysis

Note: As of June 22, 2025, shortly after the release of Hailuo-02. The Arena ELO scores are determined by responses from users in the Artificial Analysis Video Arena.

Artificial Analysis Speech Arena Leaderboard (evaluation of speech models)

Rank	Company	Model	Arena ELO
1	**MiniMax**	**Speech-02-HD**	**1,174**
2	OpenAI	TTS-1 HD	1,146
3	OpenAI	TTS-1	1,132
4	ElevenLabs	Multilingual v2	1,114
5	ElevenLabs	Turbo v2.5	1,108
6	Cartesia	Sonic English (Oct'24)	1,103
7	Kokoro	Kokoro 82M v1.0	1,078
8	Microsoft	Azure Neural	1,056
8	Amazon	Polly Long-Form	1,056
10	Google	Studio	1,039

Source: Artificial Analysis

Note: As of June 22, 2025, shortly after the release of Speech-02. The Arena ELO scores are determined by responses from users in the Artificial Analysis Video Arena.

Competitive barriers

R&D capabilities of foundation models

The competitiveness of foundation model products is fundamentally based on the underlying foundation models. Performance improvements driven by the iteration of foundation models often far outweigh enhancements made at the application layer or through product refinement. As a result, leading foundation model products nowadays are typically developed by companies with in-house foundation model R&D capabilities, while users tend to gravitate toward top-tier products that offer the best experience. Given the rapid pace of technological advancement, players in the industry must continue investing heavily in R&D to maintain performance leadership and their competitive edge.

Commercialization capabilities

Commercialization capabilities enable foundation model companies to translate research and technologies into usable products more rapidly, shortening the cycle from technological development to tangible commercial value. By strategically selecting and developing products with the greatest potential for scalable commercialization, foundation model companies can further amplify the market impact of technological breakthroughs, improve the ROI of model development, and support the long-term sustainability of ongoing research efforts.

Organizational abilities

Developing foundation models requires the integration of expertise across multiple complex domains, including advanced algorithms, large-scale model training, infrastructure optimization, and deployment efficiency. As such, companies must rely heavily on a small pool of top-tier AI talents with deep technical capabilities. To attract and retain these individuals, companies need organizational abilities — including a compelling long-term vision, research environment, capital support, and a culture that fosters innovation and ownership. These organizational qualities form a critical barrier to entry, enabling leading players to continuously overcome technical bottlenecks and maintain a sustainable competitive advantage.

KEY COSTS AND TRENDS OF THE GLOBAL FOUNDATION MODEL MARKET

Inference cost is the major cost for companies engaged in the global foundation model market. It refers to the computational expense incurred each time a user query is processed by the model, and is typically charged on a per-token basis. With the continued maturation of foundation model technologies and increasing economies of scale in commercialization, inference costs are expected to decline significantly. According to CIC, the industry average inference cost declined from approximately US$20 per million tokens by the end of 2022 to below US$0.1 per million tokens by the end of 2024, and is expected to further decline at a approximate rate of 10 times per year.

SOURCE OF INFORMATION

CIC was commissioned to conduct research and analysis of, and produce a report on the global foundation model industry at a fee of US$115,000. The commissioned report has been prepared by CIC independently without the influence from the Company or other interested parties. CIC offers industry consulting services, commercial due diligence, and strategic consulting. With a consultant team actively tracking the latest market trends in various industries such as TMT, consumer goods and services, agriculture, chemicals, marketing and advertising, culture and entertainment, energy and industry, finance and services, healthcare, and transportation, CIC possesses the most relevant and insightful market intelligence in these sectors. CIC undertook both primary and secondary research using a variety of resources. Primary research involved interviewing key industry experts and leading industry participants. Secondary research involved analyzing data from various publicly available data sources, including annual reports published by relevant industry participants, industry associations, CIC's own internal database, etc.

The market projections in the commissioned report are based on the following key assumptions: (i) the overall global social, economic, and political environment is expected to maintain a stable trend during the forecast period, (ii) key industry drivers are likely to continue to drive market growth during the forecast period, and (iii) there is no extreme force majeure or unforeseen industry regulations in which the market may be affected either dramatically or fundamentally during the forecast period. Except as otherwise noted, all of the data and forecasts contained in this section are derived from the CIC Report.