

2021

China AI Development Platform Market Report



Instruction

Sullivan hereby releases the '2021 China AI Development Platform Market Report' annual report of the China AI series of reports. This market report takes the development platform products in the field of artificial intelligence as the core research object, and the research cycle covers the whole year of 2020. This research project will focus on sorting out information about market trends, cutting-edge technologies, business models, and competitive trends in the field of artificial intelligence, and make speculations or predictions about market development prospects from dimensions such as application scenario expansion and ecological value creation.

This research project aims to sort out the artificial intelligence development platform service system, gain insights into user characteristics, market stock space and incremental space. It also studies and determines the position of various competitors in the artificial intelligence development platform industry based on the market development prospects.

The results of this study will use the growth index to reflect the ability of competitors to maintain their current market position, and the innovation index to reflect the ability of competitors to further improve their market position.

The data in all figures, tables, and texts in this report are from Frost & Sullivan Consulting (China) and LeadLeo Research Institute surveys. The data are rounded to one decimal place.

Any content provided in the report (including but not limited to data, text, charts, images, etc.) is the exclusive and highly confidential document of LeadLeo Research Institute (unless the source is otherwise indicated in the report). Without the prior written permission of LeadLeo Research Institute, no one is allowed to copy, reproduce, disseminate, publish, quote, adapt or compile the contents of this report in any way. If any behaviour violating the above agreement occurs, LeadLeo Research Institute reserves the right to take legal measures and hold relevant personnel responsible. LeadLeo Research Institute uses "LeadLeo Research Institute" or "LeadLeo" trade name or trademark in all business activities conducted by LeadLeo Research Institute. LeadLeo Research Institute neither has other branches other than the aforementioned name nor does it authorize or employ any other third party to carry out business activities on behalf of LeadLeo Research Institute.

- ❑ **Self-developed AI chips, cloud native architecture, flexible distributed training services, and MLOps capabilities have become the core capabilities of the platform**

AI chips will continue the trend of architectural innovation, morphological evolution, and integration of software and hardware. Cloud-native applications can provide users (developers) of AI development platforms with more agile and high-quality application delivery and simpler and more efficient application management. Distributed training can provide flexible configuration of the underlying resources and improve the resource utilization of the system. MLOps brings flexibility and speed to the AI development platform.

- ❑ **Developer traffic and platform scale are the two decisive elements of AI development platform's revenue**

The business model of AI development platform is relatively simple. It profits by providing companies or developers with AI technology interfaces or AI development tools. The billing methods mainly include free mode, billing based on call volume and annual or monthly subscription.

- ❑ **Model call business's revenue will increase**

From 2016 to 2020, the revenue of China's AI development platform has expanded rapidly. In 2020, the revenue of China's AI development platform will exceed 20 billion RMB. At this stage, its four businesses, namely, computing power, data, model invocation, and deployment/maintenance, account for about 4:3:2:1 in the total revenue of the AI development platform. In the future, as the proportion of inferred applications increases, the proportion of data business revenue is expected to decline. With the deepening of the application of AI in various vertical scenarios, the proportion of model call business revenue is expected to increase.

Table of Contents

◆ AI Development Platform Structure	-----	05
◆ AI Development Platform Business Mode	-----	19
◆ AI Development Platform Market Size	-----	21
◆ AI Development Platform Competitive Factors	-----	22
◆ AI Development Platform Overall Performance	-----	30
◆ AI Development Platform Scoring Dimensions	-----	31
◆ AI Leading Companies' Structures and Case Studies	-----	33
• Amazon Web Services	-----	33
• Huawei ModelArts	-----	35
• Baidu PaddlePaddle	-----	37
• Tencent Cloud TI Platform	-----	39
◆ Terms	-----	40
◆ Methodology	-----	41
◆ Legal Disclaimer	-----	42

图表目录

◆ AI development platform structure	-----	06
◆ The whole stack integration of AI development platform	-----	09
◆ The cloud native concept closed loop	-----	09
◆ Principles of image preprocessing technology	-----	15
◆ Principles of text detection technology	-----	15
◆ Principles of text recognition technology	-----	15
◆ MLOps concept: MLOps=ML+DevOps	-----	17
◆ AI development platform business mode	-----	20
◆ AI development platform market size (calculated by revenue) , 2016-2025 estimation	-----	21
◆ Generative Adversarial Networks (GANs) algorithm flow chart	-----	24
◆ Pretraining model conceptual graph	-----	26
◆ Comprehensive competitive performance of AI development platform in China——Frost Radar ™	-----	30
◆ AI Development Platform in China: Scoring dimension——Innovation index	-----	31
◆ AI Development Platform in China: Scoring dimension——Growth index	-----	32
◆ Amazon Web Services AI Development Platform Case	-----	33
◆ Huawei ModelArts Platform Case	-----	35
◆ Baidu PaddlePaddle Platform Case	-----	37
◆ Tencent Cloud TI Platform Case	-----	39



Chapter 1

Platform Structure

AI Chips, Cloud Native Structure, Elastic Training and MLOps are becoming key indicators

Platform Structure

“Self-developed AI chips, cloud native architecture, flexible distributed training services, and MLOps capabilities have become the core evaluation indicators of the platform”

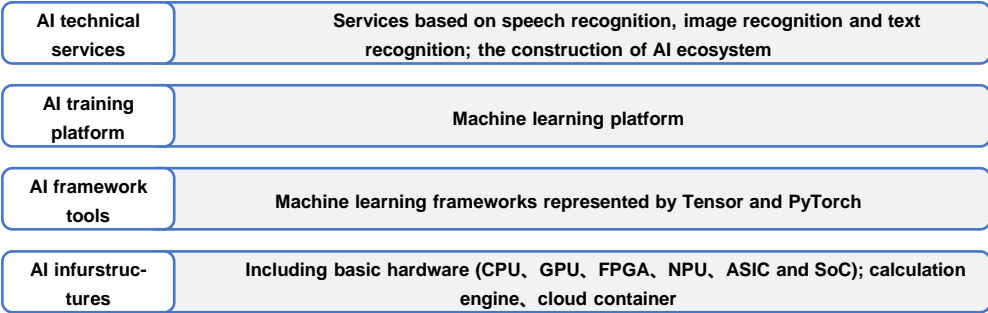
- ❑ Artificial intelligence development platform is a platform that integrates AI algorithms, computing power and development tools, open machine learning, deep learning and training model. It also provides computing power to support the development procedure. Through the form of interface invocation, developers can efficiently realize AI product development or AI empowerment.
- ❑ The AI open platform provides developers with tools and frameworks such as AI datasets, AI models and computing power to help reduce development costs. Developers can use the platform's data sets to train their models, or use the platform's algorithmic framework to customize functions they need.
- ❑ AI development platform architecture can be divided into four layers from bottom to top: infrastructure, framework tools, training platform and technical services.

1. Infrastructure: Self-developed AI chips are the core competitiveness of enterprises. Self-developed chips have the trend of framework innovation, morphological evolution and integration of both software and hardware

1.1 Basic hardware

- ❑ Currently, the mainstream AI processor is essentially a system-on-chip (SoC), which can be mainly applied in scenes related to image, video, speech and word processing. The main components of an AI processor include a specialized computing unit, a large storage unit, and a corresponding control unit. By developing their own AI chips, enterprises can adapt the chip circuit framework to their own algorithms to maximize the efficiency of computing. Self-developing AI chips will gradually become one of the core competitiveness of AI development platform enterprises.

AI development platform structure



1.1.1 AI Chip architecture innovation

- ❑ Cloud AI chips are mainly used in AI training scenarios, and computing power is one of its core metrics. In order to adapt to the applications and algorithms that need to be used in AI training, AI development platform suppliers need to develop domain-specific architecture (DSA) chips to innovate the architecture and achieve the goal of chip performance optimization. Take Huawei's Ascend chip as an example. Huawei uses the Da Vinci architecture to enhance the computing power of AI chips. Among them, the computing unit, as one of its three major components (calculation, storage and control), can perform scalar, vector and matrix operations. Huawei has deeply optimized the matrix operation in the Da Vinci architecture and customized the corresponding matrix calculation unit to support high-throughput matrix processing. This is embodied in that the Ascend chip can complete the multiplication operation of two 16*16 matrices with one instruction.
- ❑ In order to solve the problem that the existing memory access speed is seriously lagging behind the computing speed of the processor, new fully programmable, reconfigurable architecture (CGRA) chips, memory computing chips, and a new processor architecture IPU or IPU with high storage bandwidth may will be introduced into the underlying ecology of AI chips.
- ❑ In addition, chip programming methods and software architecture design will also become an important link in AI chip innovation. For example, NVIDIA uses its CUDA framework to greatly reduce the programming difficulty of its GPU, allowing GPUs to be widely used in AI acceleration. In the future, more AI processors will provide multi-layer software stacks and development tool chains to help developers use the underlying hardware resources more effectively, improve development efficiency, and reduce the defects of low flexibility of dedicated chips through the diversity of software.

1.1.2 AI Chip morphology evolution

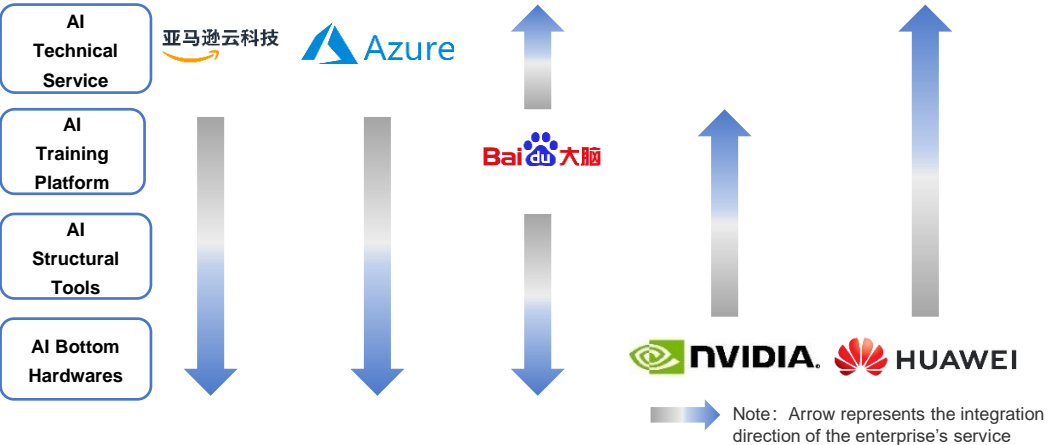
- ❑ One of the AI chip innovation goals is to maintain the high energy efficiency ratio of the chip while adapting to the evolution of AI algorithms. In the future, the system-on-chip form of general-purpose and dedicated chips will become the mainstream (CPU+NPU, CPU+ASIC, etc.) and have a wider range of applications.

- ❑ The traditional processor instruction set (including x86 and ARM, etc.) is constantly evolving in order to perform general-purpose calculations. Its basic operations are arithmetic operations (addition, subtraction, multiplication, and division) and logical operations (and or not), often requiring hundreds of instructions to complete the in-depth learning procedure of a neuron's processing. The processing efficiency of deep learning is not high. In order to solve the second pain point, the chip shape needs to break the traditional Von Neumann architecture. The neural network processor NPU uses circuits to simulate the structure of human neurons and synapses. In the NPU, the integration of storage and processing in the neural network is embodied by synaptic weights. For example, the DianNaoYu, the world's first deep learning processor instruction set proposed by the Cambricon, can directly face the processing of large-scale neurons and synapses. A group of neurons can be processed with one instruction, and the processing of neurons and synapses can be completed. The transmission of data on the chip provides a series of specialized support. In the cloud application, the Alibaba Cloud server AN1 equipped with the T-Head Hanguang NPU, in the inference application of the ResNet50 model, the Hanguang NPU can process up to 78,000 IPS images per second, which is double the performance of similar processors.

1.1.3 AI Chip software and hardware integration

- ❑ The software tools surrounding AI chips began to shift from basic computing to scene computing. In the past, chip companies represented by NVIDIA continued to build a multi-level basic software tool ecosystem such as high-performance operator libraries, communication algorithms, and inference acceleration engines centered on the CUDA programming model. At this stage, the leading AI chip companies have begun to build an integrated software and hardware platform for differentiated scenarios. The business model has expanded from the provision of hardware support services to the provision of technical production tools and technical services, and the realization of underlying chips, programming frameworks, industry algorithm libraries, and detailed full-stack and efficient integration of R&D platforms by scenes to cultivate a computing ecology for diversified industry scenes and seize market segments. At the same time, companies can also provide modular services according to customer needs, provide customers with less capable services, and enhance the degree of customization of services.

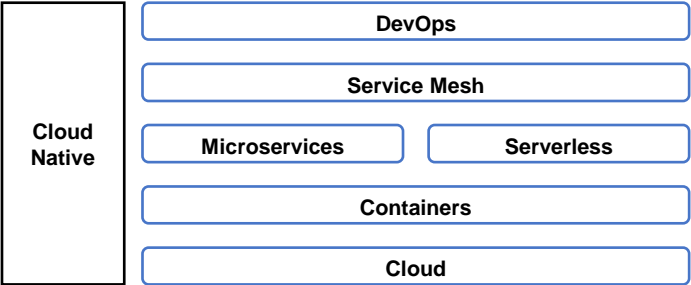
The whole integration of AI development platform



1.2 Cloud Native

- Cloud native technologies enable organizations to build and run applications that can scale flexibly in new dynamic environments such as public, private and hybrid clouds. Cloud native technologies include containers, service grids, microservices, immutable infrastructure, and declarative apis. These technologies enable the construction of loosely-coupled systems that are fault-tolerant, easy to manage, and easy to observe. Combined with reliable automation, cloud native technology enables engineers to easily make frequent and predictable significant changes to systems. At the infrastructure level, containers decouple application and technical architecture resources between cloud infrastructure and applications. At the application level, users can choose microservice architecture or no server architecture based on the scenario. In complex architecture scenarios, the communication of service components is controlled by service grid. Finally, the system was updated iteratively through DevOps.

The cloud native concept closed loop



- **Improvements in infrastructure:** The deep learning training platform based on cloud native can achieve complete container deployment and use, and provides elastic expansion of resources based on Kubernetes (K8s), scheduling and allocation of resources under different tasks, and is compatible with a variety of cpus and GPU processors downward. Therefore, the AI development platform based on cloud native can be quickly adapted to appropriate cloud native resources, no matter for trainings of large-scale sparse data or training of perception-oriented scenarios. For example, Alibaba Cloud PAI can provide a kernel that supports nearly linear acceleration, which enables training tasks to achieve performance enhancement and performance acceleration on a variety of engines.
- **Improvements in training links:** Cloud-native container architecture can flexibly allocate computing power resources of machine learning training and reduce cost and increase efficiency for AI development through elastic training. AI development platform's cloud can monitor the computing power of the resource pool in real time and allocate idle computing resources to the task in training when there are idle computing resources, so as to improve the computing power of the task and make the training operation converge quickly. After the task is submitted, the elastic training scheme can recycle the resources and allocate them to a new machine learning training task according to the usage of free resources in the resource pool and elastic jobs, which can ensure the computing power of the new machine learning training.
- **Improvement of user experience:** Cloud native applications can provide users (developers) with more agile and high-quality application delivery, simpler and efficient application management, faster response to business needs, and better user experience. Based on cloud native, the AI develop platform is a perfect fit for the team's online AI collaborative development, online AI teaching and the local migration to the upper cloud of AI research and development.

2. Structural Tools: Lead by Google TensorFlow and Facebook PyTorch

- At the initial stage of AI development platform construction, it is necessary to build the underlying technical framework, mainly referring to the deep learning framework. When building the underlying framework of the platform, platform operators can choose independent research and development or use external framework, both of which have advantages and disadvantages. Due to the high technical threshold of independent research and development, most manufacturers use external open source framework.
 - **Independent research and development:** The advantage of independent research and development is that the platform will not be subject to ecological constraints. Take Google's TensorFlow as an example. If platform operators use TensorFlow as the underlying framework of deep learning, their hardware APIS will only be connected to TensorFlow, which is deployed on Google Cloud. That makes the platform dependent on Google's ecosystem. The self-developed deep learning framework will give platform operators more freedom to play and reduce their dependence on external ecology. However, the threshold of framework development is high, the cycle is long, and the cost is high. Take Baidu for example, baidu established the Deep Learning Research Institute in 2013, during which a large number of relevant scientists and engineers participated in the research and development, and it took three years to release the PaddlePaddle framework of deep learning.
 - **Using external structures:** The main advantage of using external frameworks is that most external frameworks are open source and can be used directly by platform operators, which can effectively reduce the cost of platform construction and shorten the development cycle. The extra time and cost saved can be used for the development of its supporting tools. The disadvantage is that the use of external frameworks needs to rely on external ecology, which is not conducive to the construction of the platform's own ecology.
- Over 90% of the global deep learning framework is occupied by TensorFlow developed by Google and Pytorch developed by Facebook:
 - TensorFlow is the most popular deep learning framework currently. It has features such as visualization, powerful performance, and multi-purpose. TensorFlow comes with a tensorboard visualization tool that allows users to monitor and observe the training process in real time. It also supports multi-GPU and distributed training and has strong cross-platform operation capabilities. TensorFlow has multiple uses not limited to deep learning. It also has tools that support reinforcement learning and other algorithms.

- PyTorch is open sourced by Facebook, with features such as simplicity, ease of use, and detail. PyTorch has less abstraction and more intuitive design, the modeling process is simple and transparent, what you think is what you get, the code is easy to understand, and it can provide users with more details about the implementation of deep learning, such as backpropagation and other training processes. PyTorch has a more active community that can provide developers with complete documentation and guides for users to communicate and ask questions. But this community is smaller than the Tensorflow community.
- Other typical frameworks include Keras (open sourced by Google engineers), mxnet (open sourced by Amazon), PaddlePaddle (open sourced by Baidu), theano (open sourced by the University of Montreal), CNTK (open sourced by Microsoft). Among them, CNTK, Japanese startups preferred networks, Chainer, Theano and other early popular frameworks have left the race by merging with mainstream frameworks or directly stopping updates.
- The competitive landscape of the AI development platform framework has gradually become clearer: TensorFlow continues to rank first, relying on the deployment advantages of the industry. TensorFlow is more than 3 times concerned by the market compared to the PyTorch, who ranked second. After merging with Caffe2, PyTorch has greatly increased its application prevalence due to its ease of use. It has been used for more than 50% among the top academic conference papers.
- China is also rapidly adopting a systematic layout of open source development frameworks. Representative projects include Baidu PaddlePaddle, Megvii MegEngine, Huawei MindSpore, and Tsinghua University Jittor. Baidu PaddlePaddle is the first to be launched, and it has been initially applied in industrial, agricultural, and service industries. Its application depth is gradually improving. Baidu PaddlePaddle has more than 2.3 million developers, which makes it to be the largest open-source development framework in China.

3. Training platform: lower costs and increase efficiency through flexibly distributed training

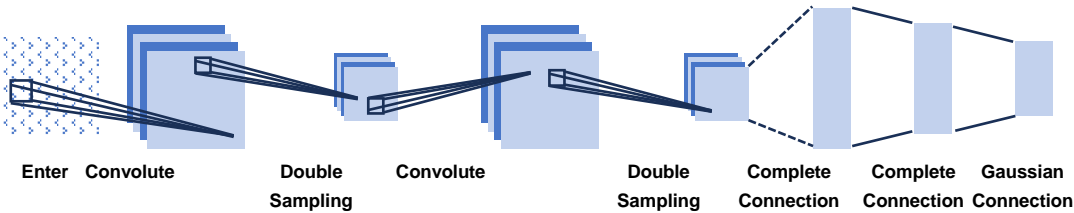
3.1 Resource allocation

- ❑ According to the fitting of real data, the amount of AI computation increases at least 10 times every year, which exceeds Moore's Law by 18 months. Therefore, the ability to adjust task resources in deep learning training becomes particularly important. With the expansion of cluster scale currently, the possibility of machine breakdown at a given moment in the cluster is increasing.
- ❑ With the increase of training model's complexity, training resources and training time increase significantly while the fault tolerance of the task decreases. In addition, as the cluster scale increases, the waste of idle resources becomes hard to ignore, and the demand for a more flexible allocation of cluster resource is increasing constantly.
- ❑ Distributed training provides elastic allocation of underlying resources. It also improves system resource utilization rate. For instance, Baidu PaddlePaddle can segment tasks through its general heterogeneous parameter server. Consequently, users can deploy distributed training tasks in heterogeneous hardware clusters and realize more efficient utilization of chips with different computing power. This provides users with training capacities which have higher throughputs and lower resource consumption.
- ❑ However, the application of distributed training also has many obstacles. To meet the huge workload may flexible training generate, developers should be able flexibly control elements in each framework and ensure the suitability of corresponding systems. In addition, if different frameworks have their own flexible training solutions, the integration of different frameworks in the AI development platform will generate high maintenance costs.
- ❑ Elastic distributed training is the trend of AI development platform services, which can reduce costs and increase efficiency for users. When users need a large number of computing resources, it can expand capacity, improve computing power and stability, and reduce model training time. When the user's computing demand is small, it reduces the underlying resource configuration and the service costs caused by resource occupation.

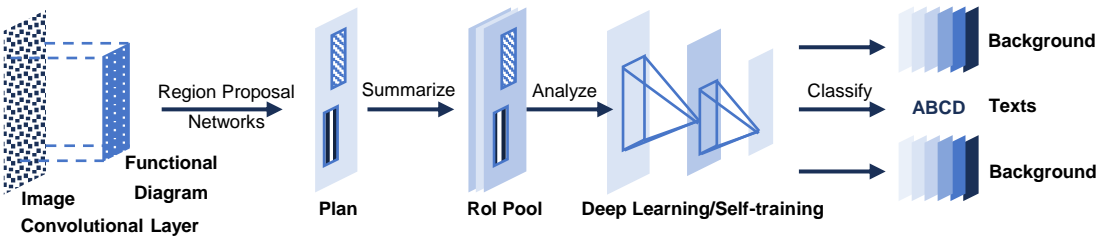
3.2 Algorithm upgrades

- ❑ Algorithm is the link between AI and big data. Internet applications such as social media, location technology, and search engines generate and store large amounts of data in real time. On the basis of massive amounts of data, AI continues to infer the interests, preferences and needs of users, generates different user portraits, and realizes the entire personalized and precise customization of digital culture from production, dissemination to acceptance.
- ❑ At this stage, the AI training platform has integrated or will integrate a variety of artificial intelligence technologies, such as computer vision, natural language processing, cross-media analysis and reasoning, intelligent adaptive learning, swarm intelligence, autonomous unmanned systems, and brain-computer interfaces, etc.:
- **Computer vision technology:** using cameras and computers instead of human eyes to identify, track and measure targets, and to perceive the environment in three dimensions.
- **Natural language processing technology:** Analyze, understand and process natural language by establishing a formal calculation model.
- **Cross-media analysis and reasoning technology:** collaborative and comprehensive processing of multiple forms, such as text, audio, video, image and other mixed and coexisting composite media objects.
- **Intelligent adaptive learning technology:** simulate the one-to-one teaching process of teachers and students and give the learning system the ability of personalized teaching.
- **Swarm intelligence technology:** the process of gathering multiple opinions into decision-making, reducing the risk of a single individual making random decisions.
- **Autonomous unmanned system technology:** A system that is operated or managed through advanced technology without manual intervention.
- **Brain-computer interface technology:** a direct connection channel established between the human or animal brain and external equipment to complete information Exchange.

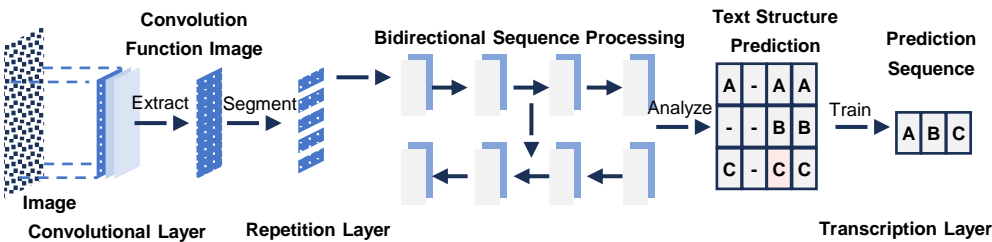
Principles of image preprocessing technology



Principles of text detection technology



Principles of text recognition technology



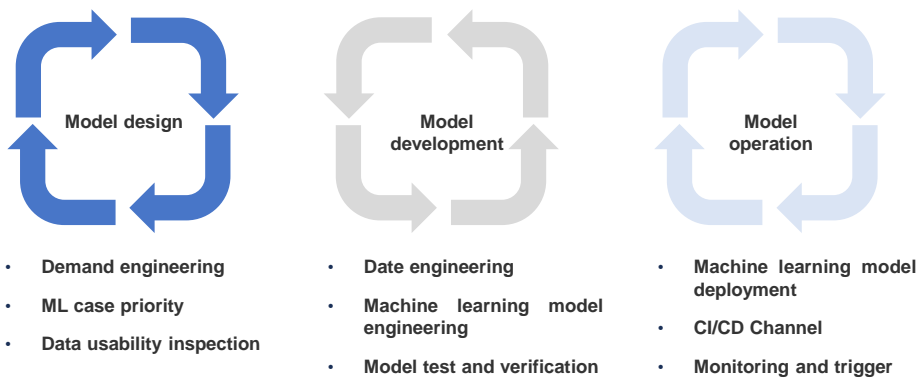
- ❑ As the prevalent application of AI learning methods in financial, medical, social and other scenarios create huge amounts of data, the AI training algorithms will be constantly trained. example, a paper of CVPR 2021 proposes a new convolutional layer named skip-convolutions, which can subtract the two frames of images and convolute only the changing part. In the image preprocessing technology, the neural network based on CNN is used as the feature extraction method, and CNN's strong learning ability can also enhance the robustness of feature extraction in the AI model. The FrameExit network, which consists of multiple cascade classifiers, can vary the number of neurons used by the model depending on the complexity of the video frames, so that when the difference between the front and back frames is large, the AI will use the whole model, and when the difference between the front and back frames is small, the AI will only use a part of the model.

4. Technical service: MLOps can enhance team efficiency

- ❑ With the development trend of industrial intelligence, AI is becoming a universal technology for transformation and upgrading of many industries. At present, the most mature and extensive application areas of AI include public security, transportation, finance, education, etc. The application requirements of AI in other industries are highly dispersed and the scenarios are also diverse, but the application requirements of AI still exist widely. Aiming at different application scenarios, the AI development platform can provide cloud-based natural language understanding, automatic speech recognition, visual search, image recognition, text-to-speech conversion, machine learning hosting and other services. The AI development platform can provide developers or business users with convenient operations for building advanced text and voice chat robots, intelligent machine learning applications, etc.
- ❑ For individual or enterprise developers, development time and development cost are the main consideration indicators for building AI applications. With the help of cloud native and flexible distributed computing architecture, users can reduce costs and increase efficiency at the level of AI model training and inference. With the help of MLOps, the team's development and deployment efficiency will be significantly improved.

- ❑ MLOps is ML's DevOps. The machine learning (ML) model built by data scientists needs to work closely with other teams (business team, engineering team, operations team, etc.). Teamwork poses challenges for communication, collaboration and coordination, and the goal of MLOps is to simplify such challenges through sound practices. MLOps brings flexibility and speed to the system: MLOps reduces development time and obtains high-quality results through reliable and effective ML lifecycle management; MLOps continues continuous development (CD), continuous integration (CI), and continuous development from DevOps. Training (CT) and other methods and tools ensure the repeatability of AI workflows and models. Developers can easily deploy high-precision machine learning models and integrated management systems to continuously monitor machine learning resources anytime and anywhere.
- ❑ MLOps also put forward higher requirements for the platform's data and hyperparameter version control, iterative development and testing, testing, security, production monitoring, infrastructure and other links. MLOps platform data plays an equally important role in defining output as written code, so the data complexity is improved compared to DevOps platform. In response to the challenges faced by the MLOps platform, the implementation process of MLOps includes five stages: use case discovery, data engineering, machine learning pipeline, production deployment, and production monitoring. Its workflow is mainly realized through agile methods.

MLOps concept: MLOps=ML+DevOps





Chapter 2

Market Size

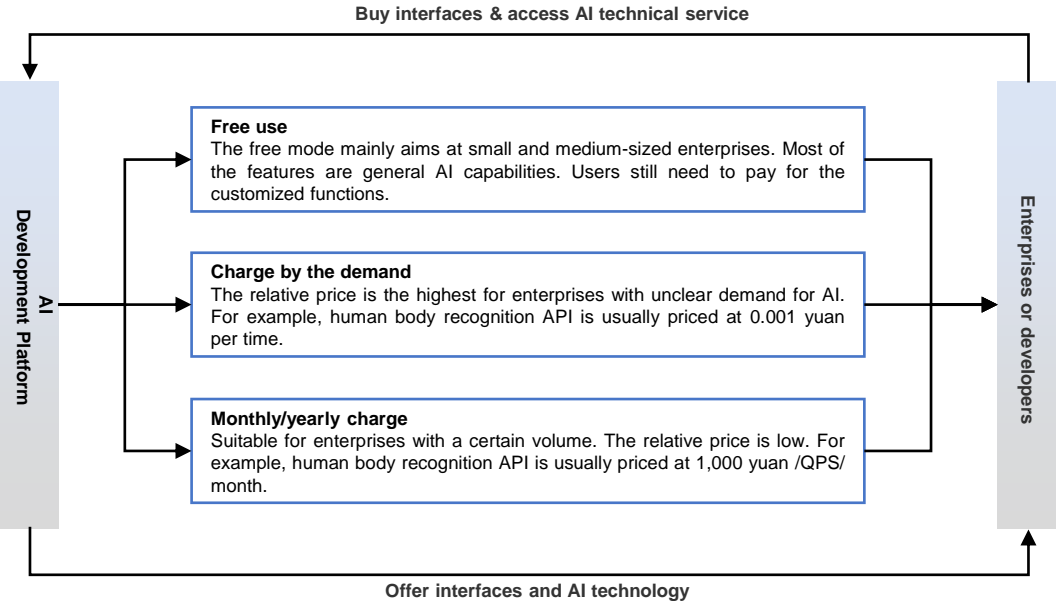
AI development platform market size exceeds 20 billion RMB, and the stock market grows steadily

**Business
Mode**

“The business model of AI development platforms is relatively simple. Developer traffic and platform size are two decisive factors for revenue.”

- ❑ AI development platform profits by providing companies or developers with AI technology interfaces or AI development tools. The billing methods mainly include free mode, billing based on pay-per-call billing, annual or monthly subscription. The free model provides companies or developers with commonly used and general AI technology interfaces such as text recognition and face recognition, with an upper limit of use, usually 1-5QPS/day. It is mainly prepared for small and medium-sized enterprises with small usage. The free model achieves profitability through data accumulation, building an AI ecosystem and providing additional services. Pay-per-call billing is also oriented to small and medium-sized enterprises. Compared with annual and monthly billing, pay-by-call billing is more expensive, which is suitable for companies that have not yet clear their demand.
- ❑ In terms of marketing, platform operators increase traffic conversion rates through free trials, subsidies, and online teaching. Large-scale platforms can further increase the conversion of traffic to users through permanent free general products. Platform operators can also explore other value-added needs of users in customer service, such as cloud services, customized AI development solutions, and so on.
- ❑ With the gradual expansion of scale, the average cost of a single customer of the AI development platform will decrease significantly, and the service profit margin will gradually increase. Therefore, achieving large-scale operations is an important development strategy for the AI development platform, which can help the platform reduce costs while also giving the platform more room for bargaining. This phenomenon also explains the underlying business logic that large manufacturers can still achieve profitability under the 'partial free' mode, and also reflects the market competitive advantages of large manufacturers compared to small and medium-sized manufacturers.

AI development platform business mode



Market Size

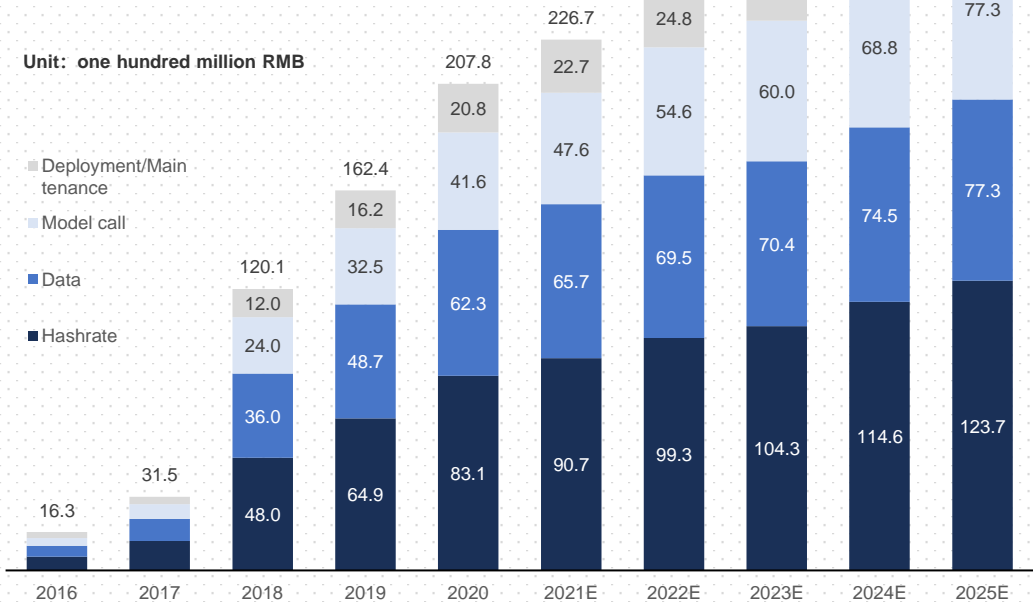
“The revenue of China's AI development platforms will exceed 20 billion yuan in 2020. The proportion of model call business revenue will increase.”

- From 2016 to 2020, the revenue scale of China's AI development platforms expanded rapidly. In 2020, the revenue of China's AI development platforms have exceeded 20 billion RMB.
- At present, computing power, data, model call and deployment/maintenance account for about 4:3:2:1 in the total revenue of the platform. The revenue share of data business is expected to decline as the share of inference applications increases. As AI becomes more widely used in vertical scenarios, the revenue share of model call business is expected to increase.

AI development platform market size (calculated by revenue) , 2016-2025 estimation

CAGR	2016-2020	2020-2025
Total	88.9%	8.3%
Hashrate	90.8%	8.3%
Date	88.3%	4.4%
Model call	69.8%	13.2%
Deployment/Maintenance	89.0%	8.3%

Unit: one hundred million RMB



Source: Frost & Sullivan



Chapter 3

Competitive factors: design a better platform for developers

Technology advances on supply side

Optimize developers' experience in the demand side

Competitive
Factors

“**The core competition of the AI development platform will focus on ‘improving its own service supply capabilities’ and ‘meeting customer needs’**”

- The users of the AI development platform are the developers of individuals or enterprises in the AI industry. The core competition of the AI development platform will revolve around how to provide developers with more efficient and convenient development platforms and other derivative services. Sullivan summarized the core competitiveness of the AI development platform as the hard power to "improve its own service supply" and the soft power to "meet customer needs".

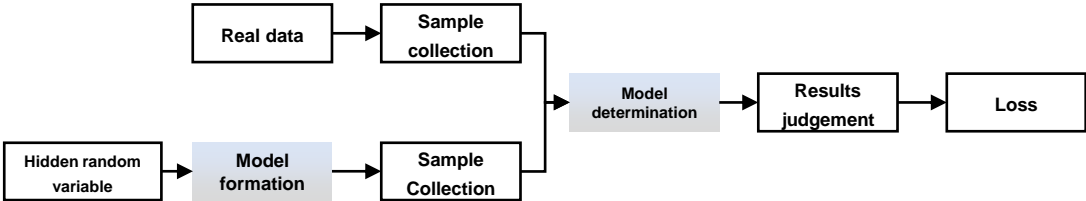
AI development platform vendors rely on the capabilities of the underlying hardware and algorithm models of the platform to provide developers with more effective AI development platform services.

Hard power 1 : Artificial Tagging—The difficult breakthrough from ‘Artificial’ to ‘Intelligence’

- The intelligent replacement of Data tagging is extremely difficult. At this stage, the tagging tool can complete basic tagging tasks with the help of algorithms, such as automatic frame recognition, automatic speech recognition, etc., and the algorithm of the tagging tool is constantly being developed and optimized.
- For AI development platforms, the intelligent labeling function is of great importance in terms of optimizing the efficiency of its own algorithms and optimizing user experience. The intelligent labeling functions that can be launched on the AI development platform include the introduction of GANs to optimize the labeling effect, the semi-supervised learning mechanism to solidify the labeling, and the introduction of a difficult case screening mechanism to optimize the labeling results and provide suggestions for improvement of data labeling based on difficult cases. However, in the actual application process, Manufacturers still need to address the limitations of the above methods.
 - GANs: Generators and discriminators need high synchronization, but in actual training it is easy to produce discriminator convergence, generator divergence scenarios, the optimization of the discriminator and generator requires extremely high design standards; GANs in the training process There will be the problem of model missing, that is, the generator function is degraded, and the same sample points are continuously generated, which makes the learning process unable to continue.
 - Semi-supervised learning: It is difficult for the model to correct its own errors; excessive smoothing may occur, causing the characteristics of the nodes to be indistinguishable.

- **Difficult case screening mechanism:** Only difficult cases can be generated during the model training process, offline difficult case mining cannot be realized, and users must adaptively modify the code to use online difficult case mining; the core idea of the difficult case screening mechanism is through bootstrapping. The method of (bootstrapping) generates a set of difficult examples, and the generation method is only judged by the loss value of the training sample during training. The evaluation dimension is single, and the improvement of model accuracy cannot be guaranteed; the algorithm idea is not mature enough to form a systematic plan.

Generative Adversarial Networks (GANs) algorithm flow chart



Hard power 2: Machine learning structure—Improve structural defects, improve users' experience, construct AI ecology

- TensorFlow and PyTorch are mainstream machine learning frameworks, with a large developer community and a large amount of mature code available. These two account for more than 90% of the global deep learning framework. But TensorFlow and PyTorch have different characteristics from each other:

TensorFlow:

- **Advantage:** It is suitable for industrial production environments, with complete solutions for model training and deployment.
- **Disadvantage:** There are many different styles of API, which are not friendly to novices. The iterative idea of distributed training is not clear. The support for cloud native is low.

PyTorch:

- **Advantage:** The programming API style is simple, intuitive and easy to understand. The built of a deep learning model based on dynamic calculation graphs can quickly debug according to stack information.
- **Disadvantage:** The deployment ecology is still in the growth stage. Some operations are not supported.

- ❑ Manufacturers such as Baidu and Huawei have launched machine learning self-developed frameworks PaddlePaddle and MindSpore.

PaddlePaddle:

- **Advantage:** Active community, complete ecological chain, user-friendly applications, full-process capability support, fast iteration pace, and support for large-scale asynchronous distributed training.
- **Disadvantage:** Most of the users are individual developers. It is not yet widely used by vendors.

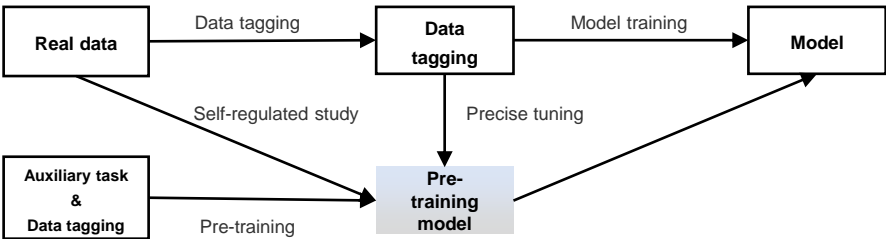
MindSpore:

- **Advantage:** Supports enhanced visualization function, differential privacy, second-order optimization algorithm, graph neural network, quantitative training, hybrid heterogeneous, MindSpore Serving, PS distributed training, MindIR, debugger, etc. Supports multiple platforms. Advocates for software and hardware collaborative design. Supports multiple modes of distributed training, etc.
 - **Disadvantage:** The number of people in the community is small. Some functions still need to be improved.
- ❑ The limited number of developers is a unified defect of the open source machine learning framework of Chinese manufacturers. There is a significant gap in the number of users compared with TensorFlow and PyTorch, and only Chinese and English are supported in language support capabilities. In contrast, TensorFlow and PyTorch will support some minority languages. So its developer ecology is more complete.
 - ❑ At present, the global machine learning framework ecology is basically stable. The common frameworks TensorFlow and PyTorch have been open sourced early, so they have ecological advantages. The self-developed framework of Chinese manufacturers can optimize the framework architecture through its technical iteration of machine learning and learning from the defects of TensorFlow and PyTorch. The research framework can provide developers with a better experience. In the long run, the developer ecology of China's self-research framework is mostly concentrated in China. In the future, more companies will use China's self-research ecology for machine learning development. However, the global machine learning framework is still expected to be led by TensorFlow and PyTorch.

Hard power 3: Pre-training model——Both ‘large’ and ‘small’

- In the future, AI development platform vendors will release multiple pre-training models that support computer vision, speech recognition, speech synthesis, natural language processing, machine translation, intelligent recommendation, business analysis and prediction, scientific computing, multi-modal data tasks and composite tasks . The pre-training model will also develop along a multi-technology path.
 - The scale of the pre-training model will increase (‘large’): the large-scale pre-training model will include more than 100 billion parameters, and the cost of a single training is expected to exceed 10 million U.S. dollars. Therefore, the pre-training model will be equipped with training optimization techniques including mixed precision training, data parallelism, model parallelism, Lamb optimizer, three-dimensional parallel training, and sparse attention acceleration. But this kind of pre-training model process is cumbersome and can only be deployed in cloud applications.
 - The pre-trained model will improve its flexibility (‘small’) through compression and acceleration: compression based on pre-trained language models such as knowledge distillation and pruning, or compression through matrix parameter decomposition, parameter sharing, and model structure design and search. Remove the redundant part of the parameter matrix and make the model "small". Or quantization-based pre-training language model compression, reducing the bit value required for numerical representation, such as quantizing a 32-bit floating point number calculation into an 8-bit or even 4-bit floating point number, simplifying the calculation process. The compressed pre-training model can be applied to the device side, and the application value is extremely wide.
- In the future, AI development platform vendors need to continuously optimize the original training methods, accelerate the training speed for training models such as Resnet50-v1.5, SSD-ResNet34, 3D UNET, RNNT, Openpose, YOLO, BERT, DLRM, etc. Or propose new training methods and improve the maturity of the pre-training model.

Pretraining model conceptual graph



AI development platform vendors improve the developers' platform using experience by providing flexible services and simplifying the operating procedures of developers. It aims to build vendors soft competitiveness.

Soft power 1: AutoML—Lowering AI development barrier, enhancing AI development efficiency

- ❑ AutoML is one of the important trends in the field of artificial intelligence. AutoML will be able to integrate the iterative process into traditional machine learning to build an automated process and greatly reduce the threshold of machine learning: AutoML is a machine learning process that uses a series of algorithms and heuristics to achieve from data selection to modeling automation. Researchers only need to input meta-knowledge (convolution operation process/problem description, etc.), and the algorithm can automatically select the appropriate data, automatically optimize the model structure and configuration, automatically train the model and adapt it to be deployed to different devices.
- ❑ AutoML can help AI development platforms to automatically complete tasks such as neural structure search, model selection, feature engineering, hyperparameter tuning, and model compression. Classification or regression problems that rely on structured or semi-structured data can be automated through AutoML, which greatly improves the efficiency of AI training.
- ❑ However, there are still some difficulties to be solved in the development path of AutoML. Firstly, AutoML still needs a lot of computing power, and companies still need to try more solutions in the research and development process. Secondly, AutoML needs to maintain a certain degree of transparency while increasing the processing complexity to allow users of the model to confirm the quality of the model. As an automation tool, AutoML has limitations in terms of resource optimization and iteration, complex model processing and feature engineering while improving work efficiency.

Soft power 2: Developers centered—Enhancing platform service capacity to built an ecosystem

- ❑ The AI development platform is a service for developers. The ability of the platform to meet the needs of developers, improve platform compatibility, and provide developers with a better development experience should also become an important evaluation criterion.
 - **In terms of data preparation functions**, the AI development platform can provide multiple data access methods including local data set loading, third-party open-source data set loading, and cloud data set calling. The platform can also provide multiple types of data annotation service models. It can also visualize the data on the operation panel.

- **In terms of model training functions**, the AI development platform can improve the compatibility of machine learning frameworks, programming languages, and cloud IDE tools. It provides customized and modular algorithm modification methods. At this stage, mainstream AI development platforms can support model management services such as flexible training, real-time monitoring of computing resources, heterogeneous training of hardware devices, multiple parallel training modes, and pre-training model migration, providing developers with convenient AI development services.
- **In terms of model management and deployment functions**, the AI development platform research and development direction covers the provision of improving the compatibility of the AI development platform, such as supporting more programming languages, supporting CI/CD, supporting third-party AIOps tools, and supporting users to build their own workflows. Machine learning workflow construction services, support model deployment monitoring services, including model drift monitoring, resource load monitoring, automatic alarms, and visual presentation of monitoring indicators.
- **In terms of account management function service capabilities**, some mainstream AI development platforms choose to open some free resources such as computing resources, storage resources, data set resources, model resources, etc., to provide developers with platform experience services. Most AI development platforms provide multiple charging modes including payment, prepayment, subscription payment (such as annual fee and monthly fee). These solutions managed to enhance the flexibility of platform fees and meet the needs of different types of developers.



Chapter 4

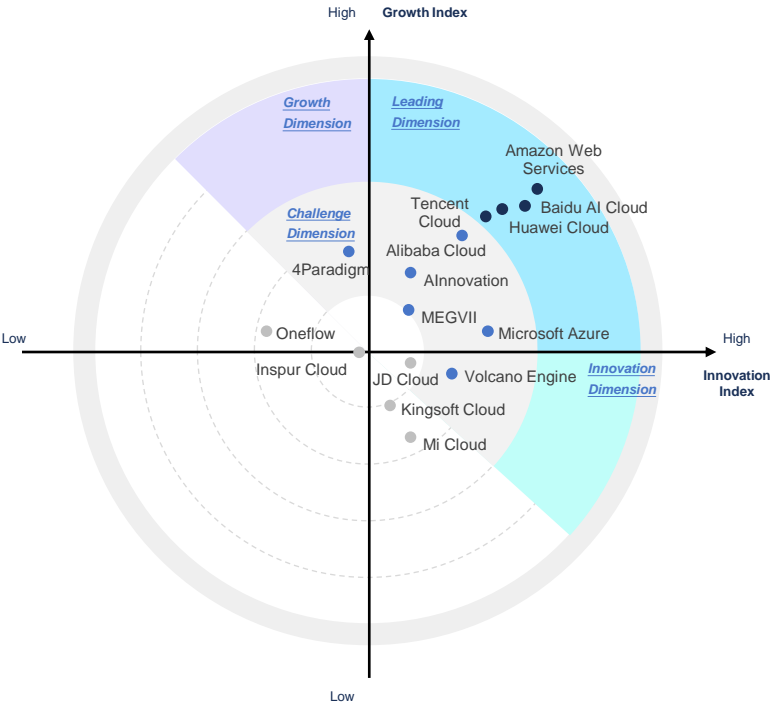
Competitive landscape

The head manufacturer has realized the full link connection, which enables technical services through technical facilities, framework tools, and training platforms

Overall
Performance

“ The AI development platform market in China is in the stage of technological maturation and platform improvements. Competitors in the AI development platform market have competitive advantages in their innovation and growth ability. ”

Comprehensive competitive performance of AI development platform in China——Frost Radar™



Note: The circle's growth from inside to outside corresponds to the comprehensive score from low to high. The competitiveness is synthesized by the 'innovation index' and 'growth index'.

The AI development platform application market in China is growing steadily. Our analysis result is only applicable to the current stage of this market.

Sullivan will continue to focus on the market of AI development platform and capture the competition trend.

- ❑ The horizontal axis represents 'innovation index':
 - It measures the innovation ability of competitors in the AI development platform market. Large abscissa represents rich service functions and strong product tuning ability.
- ❑ The vertical axis represents 'growth index':
 - It measures the competitiveness of competitors in the fields of product structure, product function and the performance growth. High ordinate represents high growth capacity of its AI development platform product.

Scoring
Dimensions

“ This report assessed and analyzed AI develop platform based on the innovation rate assessing system. Technology innovation and service innovation are two components being assessed. ”

AI Development Platform in China: Scoring dimension——Innovation index

		1 st level index	2 nd level index	Key points
Innovation Index	Technology innovation		Basic infrastructure	Assess performance in AI chips and the development and innovation of AI server etc.
			Data collection and label	Assess performance in intelligence data collection, label, analysis, filtration, enhancement etc.
			Bottom framework	Assess performance in the innovation of AI deep learning structure etc.
			Algorithm model	Assess performance of AI algorithm model in correctness, efficiency and universality of scenes etc.
	Service innovation		Cloud-native structure	Assess performance of AI develop platform in the practice of cloud-native structure, optimization of cloud computing resource and optimization of cloud develop environment etc.
			Cloud security governance	Assess performance of AI develop platform in the practice of cloud security technology and the ability of enhancing the overall platform security etc.
			Big data management	Assess performance of AI develop platform in the practice of big data management and the ability of enhancing the value of platform's database etc.
			Visual development	Assess performance of AI develop platform's ability in visualization, 0 entry barrier development functional elements, simplification of developing procedure and the ability of lowing develop barrier etc.

Source: Sullivan Data

Scoring
Dimensions

“ This report assessed and analyzed AI develop platform based on the innovation rate assessing system. Service and ecosystem are two components being assessed. ”

AI Development Platform in China: Scoring dimension——Growth index

Growth Index	1 st level index	2 nd level index	Key points
	Service	Computing service	Assess the computing performance of AI develop platform in its size, deployment and management etc.
		Data service	Assess the performance of AI develop platform in data sample size, product compatibility and product diversity etc.
		Algorithm service	Assess performance of AI develop platform in the compatibility of deep learning structure and the diversity, compatibility and portability of algorithm model product etc.
		Platform service	Assess performance of AI develop platform in data security and management, AI develop management and AI deployment etc.
		Pricing strategy	Assess performance of AI develop platform in the flexibility of service pricing and charging etc.
	Ecosystem	Ecological prosperity	Assess performance of AI develop platform in the prosperity of its development community, market influence and application range etc.
		Ecological development	Assess performance of AI develop platform in sustainable growth and the self-defense ability towards external threats etc.

Source: Sullivan Data

Amazon Web Services

“Amazon Web Services aims to put machine learning in the hands of every developer, provide AI development platform services with high compatibility and high function modularization.”

- ❑ Amazon Web Services provides extensive and in-depth machine learning services and supporting cloud infrastructure, so that developers, data scientists, and expert practitioners can easily use its machine learning technology. Amazon Web Services has become the choice of more than 100,000 customers to run their machine learning workloads.
- ❑ Amazon Web Services AI development platform has extremely high compatibility and rich application modules. Amazon Web Services Marketplace can support hundreds of algorithms and models, and its framework and toolkits are optimized for Amazon Web Services, which can help data scientists increase production efficiency by nearly 10 times. Amazon Web Services application modules include automated data extraction and analysis, business indicator analysis, voice artificial intelligence, machine vision, code, development, operation and maintenance, etc.
- ❑ Amazon Web Services provides services for different AI/ML users from three different levels:
 1. At the bottom level, for users who wish to independently develop ML capabilities, Amazon Web Services provides a broad and in-depth ML infrastructure service portfolio, including a wealth of processors and acceleration solutions for customers to choose from. For example, Amazon EC2 P4d instances can provide supercomputer-level performance for complex ML training; Amazon EC2 Inf1 instances are based on Amazon Web Services' self-developed machine learning inference chip Amazon Inferentia, which can provide high-performance machine learning inference in the cloud at low cost. At the same time, Amazon Web Services provides customers with very good openness in the choice of frameworks, supporting TensorFlow, PyTorch, Apache MXNet and all common ML frameworks; users can directly use these frameworks in Amazon SageMaker, or through Deep Learning AMI and The deep learning container uses optimized, new versions of commonly used deep learning frameworks and open-source toolkits.
 2. The middle-tier Amazon SageMaker is a broad feature set specifically built for ML by Amazon Web Services, which provides ML capabilities on a large scale in a fully managed manner for machine learning developers and data scientists. It eliminates the complexity of each step of the ML workflow, enabling customers to significantly simplify and accelerate the construction, training, deployment, and tuning of ML.
 3. The upper-level AI service can provide ready-made AI functions for users who do not have the ability to develop AIML through API calls. These services cover various functions such as text analysis, code inspection, chatbots, demand forecasting, document analysis, enterprise search, anti-fraud, image and video analysis, personalized recommendation, real-time translation, text language conversion, transcription, etc., and can be quickly developed into Various AI solutions for specific application scenarios.

Amazon
Web
Services

“Amazon Web Services aims to put machine learning in the hands of every developer, provide AI development platform services with high compatibility and high function modularization.”

Amazon Web Services AI Development Platform Case



- Through the use of Amazon Web Services, Huya realized the deployment of Huya's overseas live broadcast application Nimo TV within a few days, and the operation was extremely stable. In addition, Amazon Web Services uses GPU and self-developed AI chips to help Huya Live to provide online reasoning services at the edge.



- New Hope combines the popular 'live delivery', uses machine learning artificial intelligence services to efficiently match anchors for specific products, minimizes the application threshold, and quickly empowers through internal pilots to expand to other FMCG business areas or innovation competitions. In terms of the supply chain, New Hope uses tools such as the Internet of things, edge computing, and advanced data analysis to digitally transform the factory, feed back the factory through Amazon Web Services AI/ML services, improve production efficiency, and firmly grasp the 'two ends' of the digital transformation and upgrading of traditional enterprises.



- As it eliminates the tasks of selection, procurement, hardware operation and maintenance in traditional IT construction after the migration of New Century Healthcare to Amazon Web Services, it provides more possibilities for medical service innovation and allows IT personnel to truly participate in business innovation. In order to cooperate with medical service innovation, New Century Healthcare established a remote medical imaging center. Based on the electronic film business, a large number of medical image samples are used for machine learning through the cloud, and pediatric-specific models are trained to explore the application of AI-assisted diagnosis of pediatric medical imaging. It can effectively reduce the rate of misdiagnosis and improve diagnosis efficiency.

Huawei
Cloud

“Huawei ModelArts is a one-stop AI development platform for developers. Its application scenarios focus on intelligent recommendation, field research and product quality inspection etc.”

- Huawei ModelArts is the one-stop AI development platform for developers. It provides huge amounts of data preprocessing and semi-automatic labeling for machine learning and deep learning , large-scale distributed training, automatic model generation, and end - side - cloud model on-demand deployment capabilities. It helps users to quickly create and deploy model, workflow management cycle all AI. The core functions of Huawei ModelArts AI development platform include:
 1. **Data management:** ModelArts covers image, sound, text and video (4 categories of data formats & 9 kinds of annotation tools) while providing intelligent annotation, team annotation which can greatly improve the annotation efficiency. The platform supports data cleaning, data enhancement, data inspection and other common data processing capabilities;
 2. **Development management:** ModelArts can be developed on the cloud through the (management console) interface. It can also provide Python SDK functions, which includes creating and training models and deployment services;
 3. **Training management:** ModelArts' universal AI modeling workflow which used large model (EI-backbone) as core;
 4. **Model management:** ModelArts supports importing models from training, template, container image and OBS.
 5. **Deployment management:** ModelArts supports online reasoning, batch reasoning, and edge reasoning. Meanwhile, ModelArts' high concurrent online reasoning can meet the demand of large online business volume. Its high throughput and batch reasoning can quickly solve the demand of deposition data reasoning. Its high flexibility edge deployment enables the reasoning action to be completed in the local environment.
 6. **Mirror management:** ModelArts uses container technology at the bottom layer, which has high flexibility and is convenient to support the start-up requirements of any computing engine.

Huawei
Cloud

“Huawei ModelArts is a one-stop AI development platform for developers. Its application scenarios focus on intelligent recommendation, field research and product quality inspection etc.”

Huawei ModelArts Platform Case



- In 2018, KingMed Diagnostics and Huawei signed a cooperation agreement to cooperate in the field of artificial intelligence. The cooperation between KingMed Diagnostics and Huawei Cloud mainly focuses on the development of AI-assisted cervical cancer screening models. This development is based on the 43.5 million cervical cytology screening samples of Jinyu over the years, from which nearly 200,000 image blocks were selected for precise labeling and AI-assisted screening model training. At present, the AI model has been validated with a validation set of more than 20,000 sample data. The verification results show that the negative rejection rate of this model is 61.9%, which reduces the workload of doctors in manual reading; the accuracy of interpretation of negative films is higher than 99%; the detection rate of positive lesions exceeds 99.9%, ensuring no missed diagnosis and avoiding false negatives.



- HEXA IoT developers developed a pepper identification model based on Huawei Cloud ModelArts, unified the pepper sorting standards, effectively reduced the judgment error caused by manual screening, and increased the overall sorting efficiency by 50%. HEXA FOOD uses the Huawei Cloud ModelArts one-stop AI development platform to train the pepper recognition model and sends the model to the edge smart device Huawei Atlas 500 smart station. Through real-time analysis and judgment, the pepper is intelligently sorted, high-quality raw materials are screened, and peppers are ground. Powder is used to produce all kinds of spices, greatly improving the production efficiency and quality of spices.

Baidu

“Baidu PaddlePaddle is a comprehensive open-source AI development platform, which is committed to deeply integrate into industries. It is used widely in industrial, agricultural and service industries.”

- Based on Baidu's years of deep learning technology research and business application, PaddlePaddle integrates the core framework of deep learning, basic model library, end-to-end development suite, tool components and service platform. Officially opened in 2016, PaddlePaddle is an industry-level deep learning platform with comprehensive open source, leading technology and complete functions. PaddlePaddle originated from industrial practice and has always been committed to in-depth integration with the industry. At present, flying paddle has been widely used in industrial, agricultural, service and other scenarios, serving more than 2.65 million developers, and helping more and more industries to achieve AI empowerment together with partners. The core technologies of PaddlePaddle include:
 1. **Deep learning framework:** PaddlePaddle's deep learning framework is based on deep learning computing abstraction that is consistent with programming and the corresponding front and back-end design. It has easy-to-learn and easy-to-use front-end programming interface and has leading training performance. The internal core architecture is unified and efficient. It is easier for ordinary developers to get started.
 2. **Deep learning model training technology:** PaddlePaddle has made great breakthroughs in the training technology of super-scale deep learning model. It has realized the ability of parallel training of billions of sparse features, trillions of parameters and hundreds of nodes before other structures. PaddlePaddle has also solved the online learning and deployment difficulty in a super-scale deep learning model.
 3. **High-performance reasoning engine for multi-terminal and multi-platform deployment:** PaddlePaddle provides comprehensive supports for reasoning deployment, which can easily deploy models to different platform devices such as cloud server, mobile terminal and edge terminal. It has comprehensive leading reasoning speed and is compatible with models trained by other open-source frameworks.
 4. **Industry-level open-source model library:** PaddlePaddle has built a large-scale official model library, with a total number of more than 270 algorithms, including mainstream models honed by long-term industrial practice and international competition winning models. It provides several end-to-end development kits for semantic understanding, image classification, target detection, image segmentation, word recognition (OCR), speech synthesis and other scenarios to meet the needs of enterprises for low-cost development and rapid integration.

Baidu

“Baidu PaddlePaddle is a comprehensive open-source AI development platform, which is committed to deeply integrate into industries. It is used widely in industrial, agricultural and service industries.”

Baidu PaddlePaddle Platform Case



- ❑ COMAC's North Research Center uses the PaddlePaddle open-source deep learning platform to develop AI solutions in the non-destructive testing process of typical aircraft structures, replacing technicians in defect analysis tasks, improving the intelligence of the process, and realizing the application of AI technology. Baidu PaddleSeg can solve the above problems through data enhancement, algorithm selection and other technologies, which effectively help technicians in traditional industries to develop AI algorithms.



- ❑ OPPO has upgraded the application store based on the ability of PaddlePaddle: adopts high-performance data reading interface Dataset to complete large-scale data reading. Adopts large-scale CPU distributed fully asynchronous training model based on parameter server. Use lower resources consume to train massive amounts of data and extremely sparse parameters.



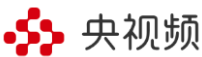
- ❑ Baidu PaddlePaddle open-source deep learning platform has officially completed the adaptation to Rockchip's AI chips RK1808 and RK1806, and is fully compatible with Paddle Lite, a lightweight inference engine for PaddlePaddle. By adapting the PaddlePaddle open-source deep learning platform, Rockchip will be able to better empower the business needs of domestic users and provide strong computing power for end-side AI.



“Tencent Cloud TI platform can provide full-stack AI development services and support a variety of learning frameworks and algorithms, which can help developers achieve convenient development”

- Tencent provides AI development solutions for a variety of business scenarios and continues to reduce the cost of building and applying AI capabilities. Tencent Cloud TI platform can link up the whole process of industry +AI landing, helping users to create and deploy AI applications quickly. Platform function modules can be decoupled, delivered and deployed corresponding to demand. Tencent Cloud TI platform includes machine learning platform TI-One, AI application service platform TI-matrix and data tagging platform TI-DataTruth.
- 1. TI-One is a machine learning platform designed for data scientists. It provides support for the whole process from data preprocessing, model training, automatic learning, model evaluation to model publishing and deployment. TI-One has rich built-in algorithm components. It supports a variety of algorithm frameworks. Drag-and-drop task flow design is novice friendly and meets the needs of a variety of AI application scenarios.
- 2. TI-Matrix AI application service platform is also designed for AI application developers. It provides fast access to models, data and intelligent devices. It also provides model services, application workflow orchestration, cloud side scheduling and other functions to quickly build intelligent applications. TI-Matrix can independently cover cloud side model service release management, AI studio and application development.
- 3. TI-DataTruth is a data tagging platform which provides intelligent data production services such as data tagging, data crowdsourcing management, and scene data mining for AI data service providers. As a data tagging platform, TI-DataTruth provides professional data tagging tool supports. The platform focuses on the two core demands of high efficiency and data security to create a data production process platform. It can quickly improve the ability of AI algorithm for enterprises and offer quality services for the implementation of AI scenes.

Tencent Cloud TI Platform Case



Media AI Medium Platform solution:

- China Media Group Mobile is a typically successful landing platform of Tencent Cloud TI Platform in large scales. central video is the first national 5G new media platform in China. TI, being the AI platform for China Media Group Mobile, connects the training sequence of data and algorithm based on visual modeling. It can support the video recommendation, content approval model and algorithm service release from side to side.



Analyst workbench solution:

- Established the 'three horizontal and two vertical' data governance system of the whole bank. Based on TI-One, the Hefei Software Center created an analyst workbench as the bearer of the analysis layer. The platform supports the implementation of AI business in the industry, helping to realize 'capital flow prediction', 'post-loan risk prediction', 'label construction' and other AI scenarios.

Terms

- ◆ **QPS (Queries per-second):** QPS is a measure of how much traffic is being processed by a particular query server in a given period of time, which can be understood as the number of concurrent requests per second. One QPS is equal to 86,400 calls approximately.
- ◆ **API (Application Programming Interface):** APIs are predefined functions designed to provide applications and developers with the ability to access a set of routines based on a piece of software or hardware without having to access the source code or understand the details of the inner workings.
- ◆ **Convolution:** The mathematical concept of generating a third function from two functions f and g , representing the integral of the overlapping length of f and g by the product of the overlapping values flipped and shifted.
- ◆ **CGRA (Coarse-grained Reconfigurable Architecture):** CGRA is a spatial parallel computing model, which organizes computing resources with different granularity and functions by spatial hardware structure. During operation, according to the characteristics of data flow, the configured hardware resources are interconnected to form a relatively fixed computing path, and the way of computing is close to "special circuit". When the algorithm and application transform, again through configuration, reconfiguration for different computational paths to perform different tasks.
- ◆ **CUDA (Compute Unified Device Architecture):** CUDA is a parallel computing platform and programming model created by NVIDIA based on Graphics Processing Units (GPUs) produced by NVIDIA.
- ◆ **DevOps:** A combination of Development and Operations, DevOps is a collective term for a group of processes, methods, and systems that facilitate communication, collaboration, and integration between Development (application/software engineering), technical Operations, and quality assurance (QA) departments.
- ◆ **Data tagging:** The process of tagging metadata such as text, video, and images. The labeled data will be used to train machine learning models.
- ◆ **Cloud native:** distributed cloud based on distributed deployment and unified management, a set of cloud technology product system based on container, microservice, DevOps and other technologies.

Methodology

- ◆ Frost & Sullivan has conducted in-depth research on the market changes of 10 major industries and 54 vertical industries in China with more than 500,000 industry research samples accumulated and more than 10,000 independent research and consulting projects completed.
- ◆ Rooted on the active economic environment in China, the research institute, starting from data management and big data fields, covers the development of the industry cycle, follows from the enterprises' establishment, development, expansion, IPO and maturation. Research analysts of the institute continuously explore and evaluate the vagaries of the industrial development model, enterprise business and operation model, Interpret the evolution of the industry from a professional perspective.
- ◆ Research institute integrates the traditional and new research methods, adopts the use of self-developed algorithms, excavates the logic behind the quantitative data with the big data across industries and diversified research methods, analyses the views behind the qualitative content, describes the present situation of the industry objectively and authentically, predicts the trend of the development of industry prospectively. Every research report includes a complete presentation of the past, present and future of the industry.
- ◆ Research institute pays close attention to the latest trends of industry development. The report content and data will be updated and optimized continuously with the development of the industry, technological innovation, changes in the competitive landscape, promulgations of policies and regulations, and in-depth market research.
- ◆ Adhering to the purpose of research with originality and tenacity, the research institute analyses the industry from the perspective of strategy and reads the industry from the perspective of execution, so as to provide worthy research reports for the report readers of each industry.

Legal Disclaimer

- ◆ The copyright of this report belongs to LeadLeo. Without written permission, no organization or individual may reproduce, reproduce, publish or quote this report in any form. If the report is to be quoted or published with the permission of LeadLeo, it should be used within the permitted scope, and the source should be given as "LeadLeo Research Institute", also the report should not be quoted, deleted or modified in any way contrary to the original intention.
- ◆ The analysts in this report are of professional research capabilities and ensure that the data in the report are from legal and compliance channels. The opinions and data analysis are based on the analysts' objective understanding of the industry. This report is not subject to any third party's instruction or influence.
- ◆ The views or information contained in this report are for reference only and do not constitute any investment recommendations. This report is issued only as permitted by the relevant laws and is issued only for information purposes and does not constitute any advertisement. If permitted by law, LeadLeo may provide or seek to provide relevant services such as investment, financing or consulting for the enterprises mentioned in the report. The value, price and investment income of the company or investment subject referred to in this report will vary from time to time.
- ◆ Some of the information in this report is derived from publicly available sources, and LeadLeo makes no warranties as to the accuracy, completeness or reliability of such information. The information, opinions and speculations contained herein only reflect the judgment of the analysts of leopard at the first date of publication of this report. The descriptions in previous reports should not be taken as the basis for future performance. At different times, the LeadLeo may issue reports and articles that are inconsistent with the information, opinions and conjectures contained herein. LeadLeo does not guarantee that the information contained in this report is kept up to date. At the same time, the information contained in this report may be modified by LeadLeo without notice, and readers should pay their own attention to the corresponding updates or modifications. Any organization or individual shall be responsible for all activities carried out by it using the data, analysis, research, part or all of the contents of this report and shall be liable for any loss or injury caused by such activities.