SMU
SINGAPORE MANAGEMENT UNIVERSITY

ANALYSIS OF INNOVATION APPLICATIONS AND VALUE DRIVEN BY AIGC TECHNOLOGY: An Early Investment Perspective in the Chinese Market

Technical Appendix

Siyuan Ma
Qifu Capital

Mengyu Wang
Singapore Management University

Tianyi Zhang
Northwestern University

SIM KEE BOON INSTITUTE FOR FINANCIAL ECONOMICS

LEE KONG CHIAN SCHOOL OF BUSINESS

Appendix 1: Machine Vision and Convolutional Neural Networks

Following the introduction of generative AI/AIGC concepts and applications, it is necessary to clarify two related technical concepts: machine vision (CV) and convolutional neural networks (CNN). Computer Vision (CV) is a discipline that studies how to convert inputs such as images, sounds, and videos into a language that computers can understand. Taking 2D images as an example, all images can be broken down into very small pixels, each of which can be converted into a number (based on grayscale) or a combination of numbers. The number represented by each pixel is the feature of that pixel, and combining these numbers forms the "language" that represents the image's characteristics, making it easy for machine learning (Figure 6).

Convolutional neural networks (CNN) are a type of artificial neural network used for deep learning, allowing AI to better understand data and solve complex problems. The core of CNNs is to extract features

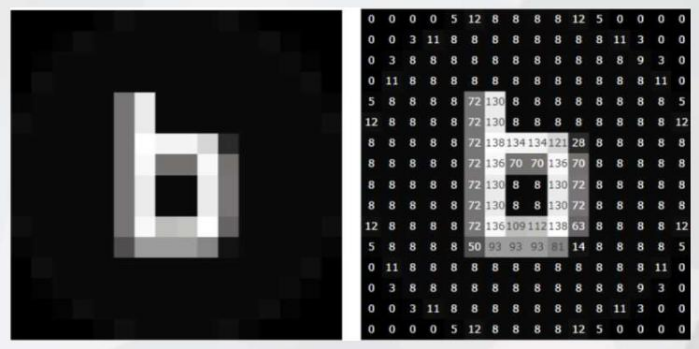
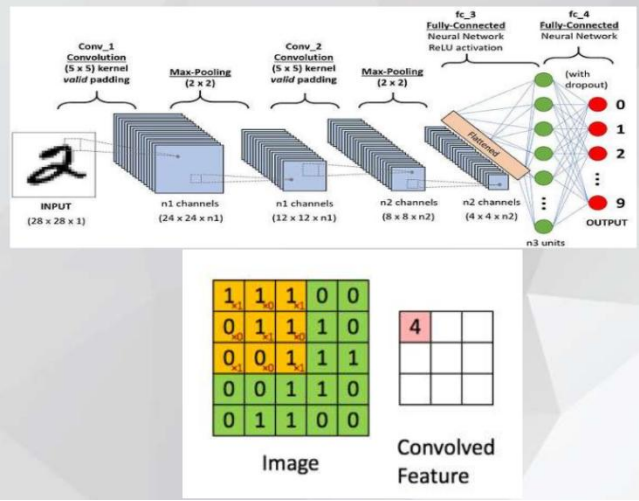


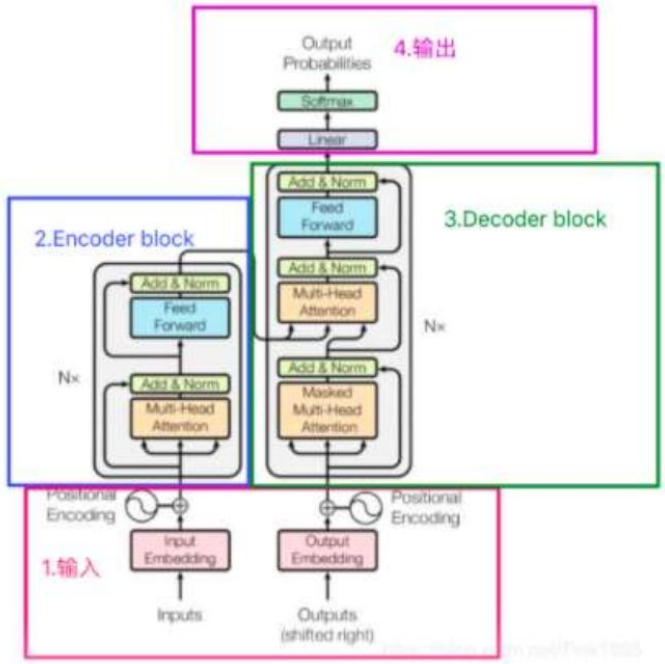
Figure 6 How machine vision understands 2D images
(Source: Internal research report compiled by Qifu Capital)

from input data and reduce neural network parameters through matrix operations, enabling more complex deep learning tasks (see Figure 7).

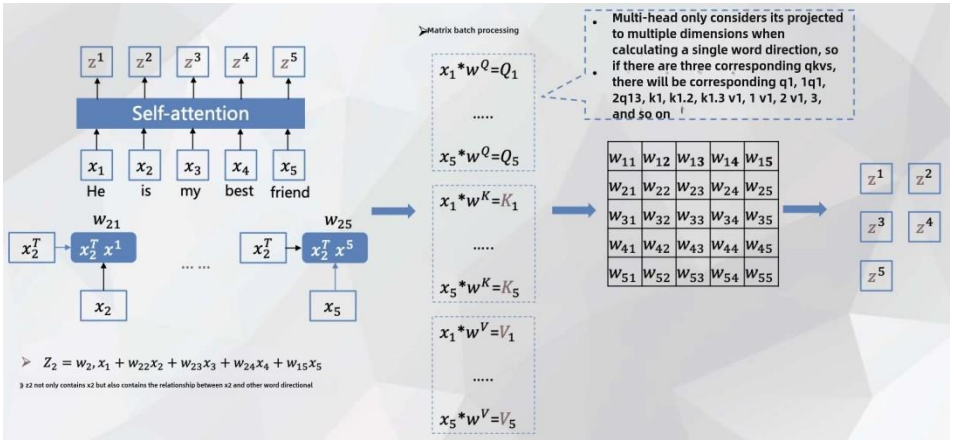


Appendix 2: Transformer Architecture & Attention Mechanism

The Transformer model is a deep learning model that uses a self-attention mechanism, mainly applied in natural language processing (NLP) and computer vision (CV) fields. Its architecture is illustrated in Figure 8, with core components including Multi-head Attention and Embedding:



- **Multi-head attention：**The core of this component is to calculate the correlation between vectors. Through matrix calculations (mapping), vectors can be transformed into Queries (Q), Keys (K), and Values (V) needed for computing attention. Self-Attention refers to a mechanism for calculating correlations, such as measuring the correlation between a word vector in an input sentence and other word vectors, including itself. Multi-head attention involves projecting vectors onto different dimensions through matrix calculations, thereby calculating the correlations between data under different missing conditions. Finally, the correlations across different dimensions are integrated to obtain the final output. The principles of Self-Attention and Multi-head Attention are shown in Figure 9.

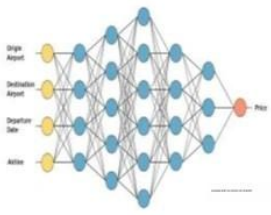
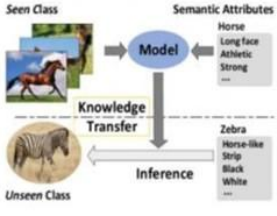
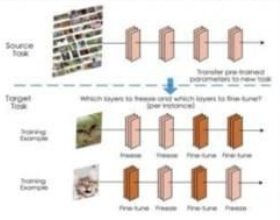


- **Embedding:** The core function of this component is to vectorize input data such as text. For example, the sentence "How are you" can be converted into a vector of pure numbers that a machine can understand, such as [0110101010], [01001010], [101001010]. During design, similar words are naturally assigned more similar vectors.

When comparing CNN and Attention mechanisms, each has its own advantages and disadvantages. CNN can be considered a simplified version of Self-attention, focusing only on information within the scanning window and having stronger relationships between adjacent elements. It performs better with smaller datasets. Self-attention considers global information, such as the relationship between every pixel in an image and all other pixels, but its early capabilities in processing images were limited by token restrictions. However, with the introduction of ViT (Vision Transformer) and Swin (Shifted Window), which compress inputs into patches, the processing capability surpassed that of CNN. In scenarios with large datasets, especially when using pre-trained models, Self-attention performs better than CNN. For multimodal models, early applications saw better image processing with CNN and better NLP processing with Transformer. The fusion of cross-modal data was challenging. With the enhanced image processing capability of Transformer's attention mechanism, multimodal large models based on Transformer architecture (e.g., image + text) have become mainstream. As Transformer architecture continues to evolve and improve, it is expected to achieve more modal integrations, such as Image-Text, Text-Audio/Speech, and other multimodal applications.

Appendix 3: Model Learning & Adjustment Methods

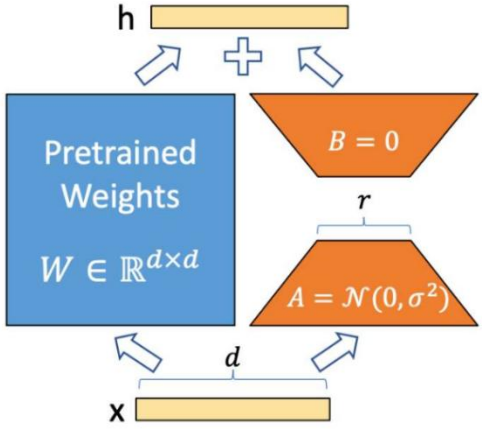
Generative AI model learning and adjustment methods are divided into four types: traditional training, few-shot learning, zero-shot learning, and fine-tuning. Each of these methods has distinct definitions and purposes, with unique advantages and disadvantages. The specific differences are detailed in Figure 10.

	Traditional training	Few-shot learning (sample book learning)	Zero-shot learning (zero sample learning)	Fine-tuning (fine adjustment)
Definition / purpose	> learn independent processing and analysis of new data through algorithm modeling and large number of data training	> the purpose of learning samples is to obtain accurate classification testing of samples with a small amount of data during training > the core is to let the machine learn to learn, understand the differences between the differences in the differences of the incidents, not a certain designated category	> the purpose of zero sample learning is to predict training data that has not appeared in the group >The difference from fla-shotlearning is that the training method is different	> personalized model training practice for specific scenes can be modified on the original pre-training large model
Illustration				
Advantages	>wide coverage and strong applicability > strong learning ability	> the cost of data collection and calculation is lower than that of traditional models and has lower data requirements for labeling >The model has high flexibility	>lower requirements for labeling data >The model has high flexibility	> can be fine-tuned based on neutral large models and applied to models in specific fields > compared with traditional training, it can reduce training time and resource consumption
Insufficient	> compared with sample books and sample learning models, the required labeling data is large > compared with sample and zero sample learning model training, the difficulty and cost are relatively large	> high requirements for examples used to train models	> it's troublesome to deal with data in multi-label classification > the zero sample learning in the field needs to be recessed and practiced/tested during transfer	> there may be overmatches > high requirements for data quality and quantity > the cost is higher than that of small sample learning and zero sample learning

Appendix 4: LoRA（Low Rank Adaptation of Large Language Models）

As large models become increasingly larger, fine-tuning by retraining all model parameters becomes increasingly difficult and expensive. LoRA can decompose the weights of each layer (scaling weights through matrix operations), reducing the number of model training parameters and improving performance and efficiency. Compared to the Adam-optimized GPT-3, which has 175 billion parameters, LoRA can reduce training parameters by 10,000 times and decrease GPU memory requirements by threefold. The principle of LoRA is shown in Figure 11.

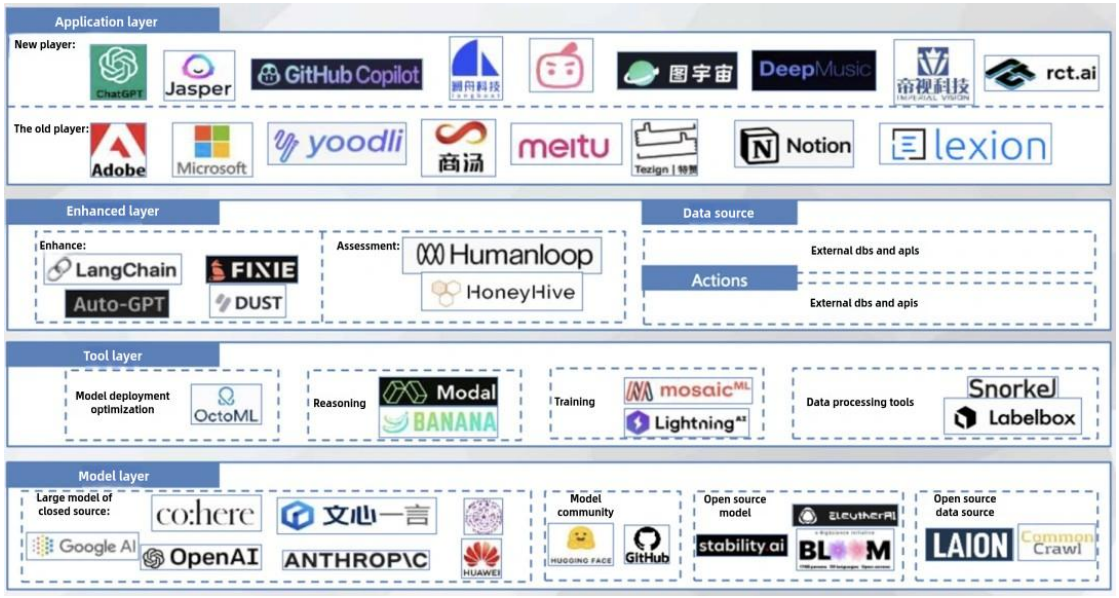
LoRA offers multiple advantages. First, by freezing the weights of pre-trained models, memory and storage usage are significantly reduced, lowering hardware requirements. Second, it achieves good results with weights several orders of magnitude smaller than the original model. Third, it allows for low-cost switching between different tasks during deployment by sharing large model parameters and only exchanging LoRA weights instead of all parameters. Future research directions for LoRA include combining it with other efficient adaptation methods and refining the mechanisms behind fine-tuning and LoRA.



Appendix 5: AI Infrastructure and AIGC Industry Overview

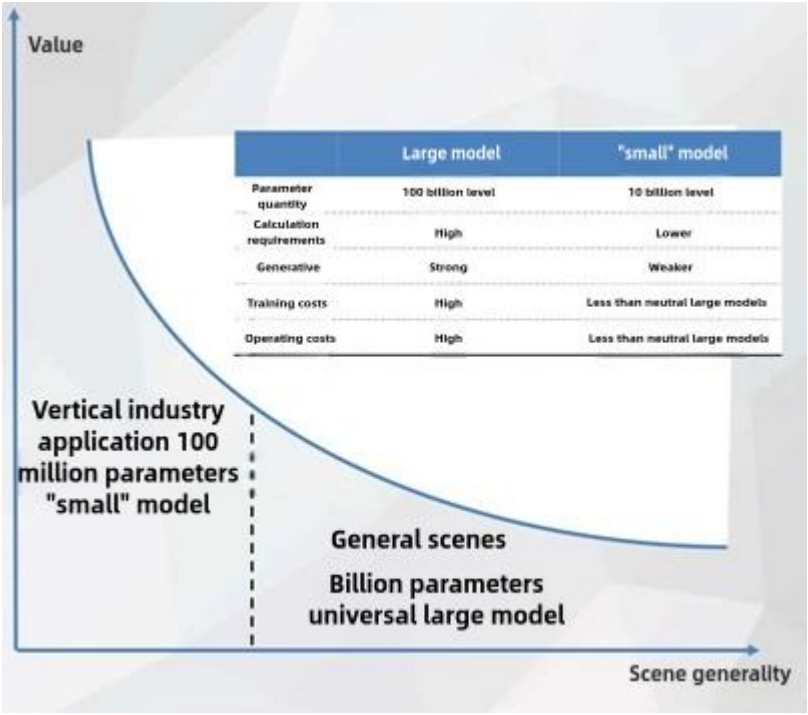
The next appendices will introduce the AI infrastructure hardware layer and provide an overview of the AIGC industry, followed by a detailed analysis of the four layers within this overview: the model layer, the tool layer, the enhancement layer, and the application layer.

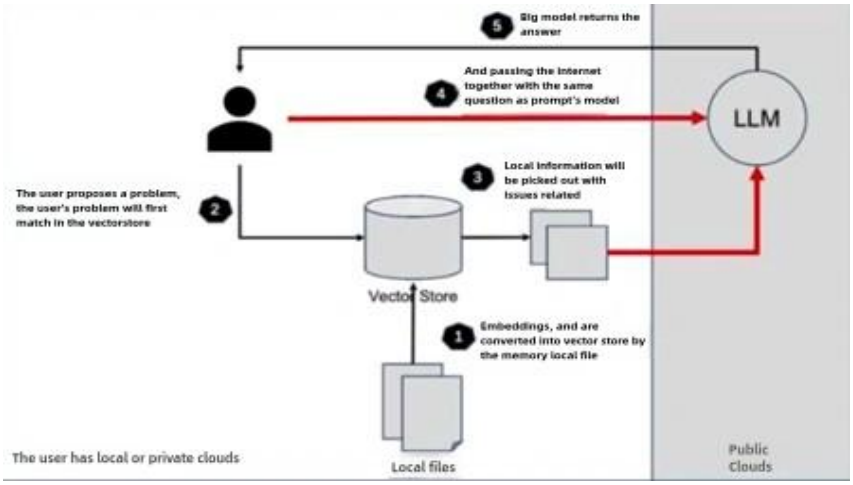
The AI infrastructure hardware layer includes CPUs, GPGPUs/AI accelerator chips (which can be divided into inference chips and training chips based on specific use cases), photonic chips, DPUs, FPGAs, switches, and GPUs. The distribution of companies involved in these areas is shown in Figure 12. The AIGC industry itself comprises four branches: the model layer, the tool layer, the enhancement layer, and the application layer, with active companies in each branch illustrated in Figure 13.



Appendix 6: Model Layer Analysis

AIGC models can be divided into large models and smaller models. The definitions and relationships between these types are shown in Figure 14. While large models offer broad applicability, they cannot fully address the high-value and complex problems in vertical industries. It is the fine-tuned industry-specific models that truly resolve enterprise pain points.

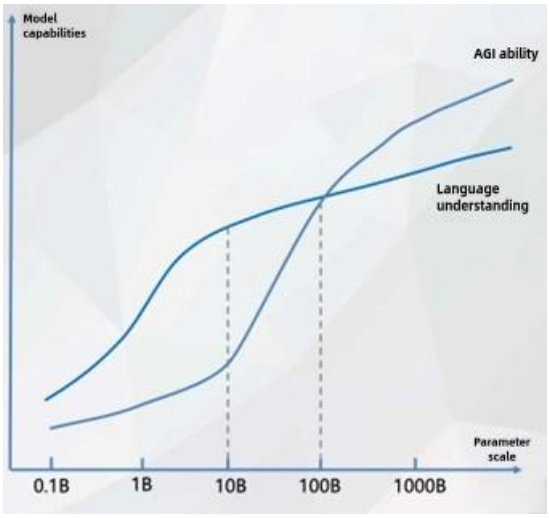




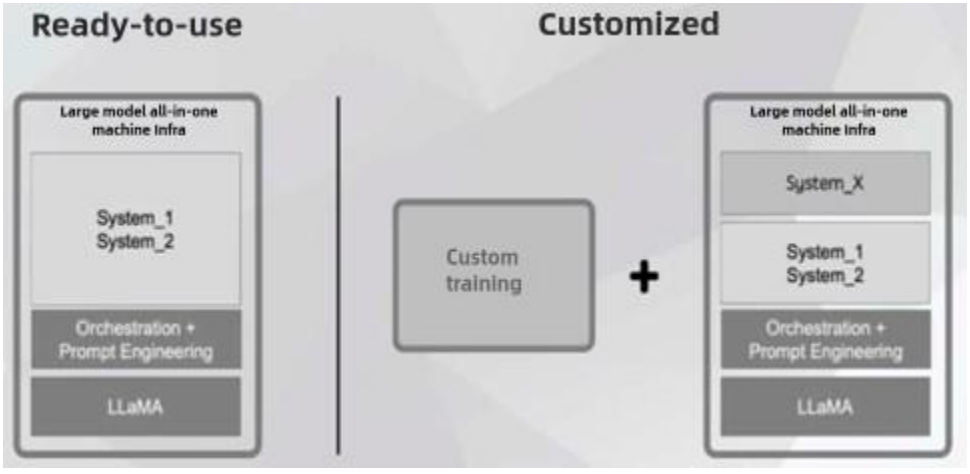
Due to data security concerns, large enterprises in China are unlikely to expose their core data assets to public large models. Connecting public large models through frameworks like Langchain, which involves packaging data assets into prompts and transmitting them remotely, carries a risk of data leakage. Hence, localized deployment of large models is necessary for data security.

Compared to large models with hundreds of billions of parameters, models with tens to hundreds of billions of parameters occupy a "rich area" where AGI capabilities and language comprehension can be effectively achieved (see Figure 16). The business model for these "smaller" models includes various systems, such as:

- **System_1:** Summarizing and sharing enterprise information. For instance, Microsoft 365 Copilot already demonstrates some OA capabilities.
- **System_2:** Analytical capabilities. Based on existing data insights, this system can be seen as an upgrade and replacement for BI tasks. Systems 1 and 2 are tasks that current large models can already accomplish.
- **System_X:** Problem-solving for complex issues. This system addresses specific problems in industries, which may involve decision-making, new product development, or complex collaboration management.

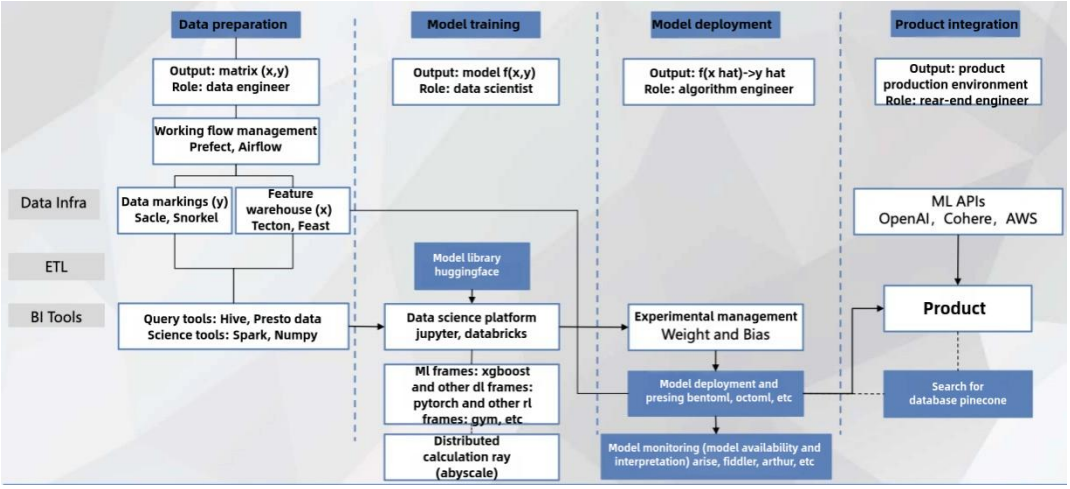


These three systems can form two business models: out-of-the-box and customized. The differences between these two models are shown in Figure 17.



Appendix 7: Tool Layer Analysis

The AI tool layer (Infra) refers to the middleware between the underlying model and the upper application layer, primarily used to assist in data preparation, model training, and deployment. Generally, the AI workflow includes data preparation, model training, model repository, model deployment, model monitoring, product integration, and vector databases (see Figure 18). Among these, the model repository, model deployment, model monitoring, and vector search databases offer significant opportunities in terms of necessity, user retention, and commercial potential.



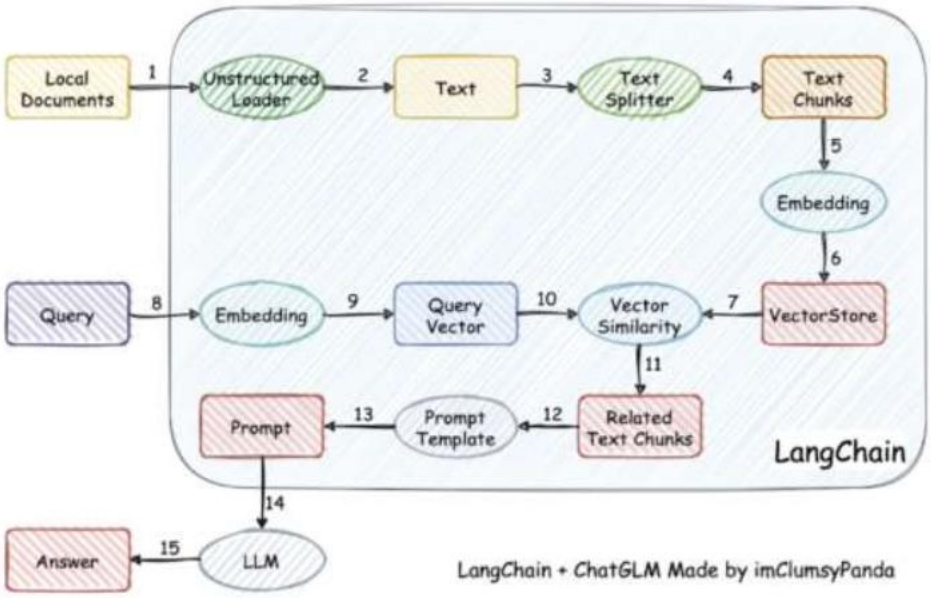
Current LLMs face three major challenges: lack of real-time information in pre-trained models, absence of vertical and enterprise-specific data, and limitations in prompt size, which hinder processing of large texts. Addressing these challenges, whether by building custom large models or fine-tuning existing ones, involves high costs for enterprises. Hence, integrating vector databases with LLMs is a viable solution. In both neural networks and deep learning, vectors are fundamental elements of information exchange. These new types of data need storage, leading to the development of vector databases. Vector databases typically use vector index-based storage methods, mapping vector data into high-dimensional spaces and constructing index structures in these spaces to support efficient similarity queries, unlike traditional relational databases.

The combination of vector databases and LLMs can achieve several functionalities, including:

- Supplementing real-time information.
- Incorporating vertical domain or company-specific data.
- Exposing hidden information within different modal data through the vectorization process, thereby supporting multimodal applications.
- Enhancing efficiency, enabling LLMs to read large texts.

With vector database + LLM, it is possible to find answers within private data domains. User queries first enter the vector database, where the vector database's similarity search capabilities precisely locate relevant data. This information is then compiled into a prompt. The compiled prompt is sent to the LLM, which

returns an answer to the user.



In terms of data security, OpenAI in the United States has already guaranteed that prompts can only be used as input, cannot be stored, and cannot be used for training. This legal assurance ensures data security. Future legal and regulatory frameworks in China concerning LLMs will also be further developed.

Appendix 8: Enhancement Layer Analysis

The enhancement layer primarily features three products: Langchain, Dust, and Fixie.

Langchain is a developer framework used for building applications driven by large language models (LLMs). Developers can directly call the encapsulated logic and code in Langchain. Its architecture and functionalities, along with supported integrations, are shown in Figure 19 and Table 1. The critical infrastructure that empowers LLMs with context and action capabilities includes:

- Prompt management, optimization, and serialization.
- Long-term memory access.
- Integration with other large models, external APIs, and databases.
- Development of agents and provision of templates.

Regarding developer tool integration, Langchain currently supports most existing developer tools and libraries. It has a rapidly growing open-source community.

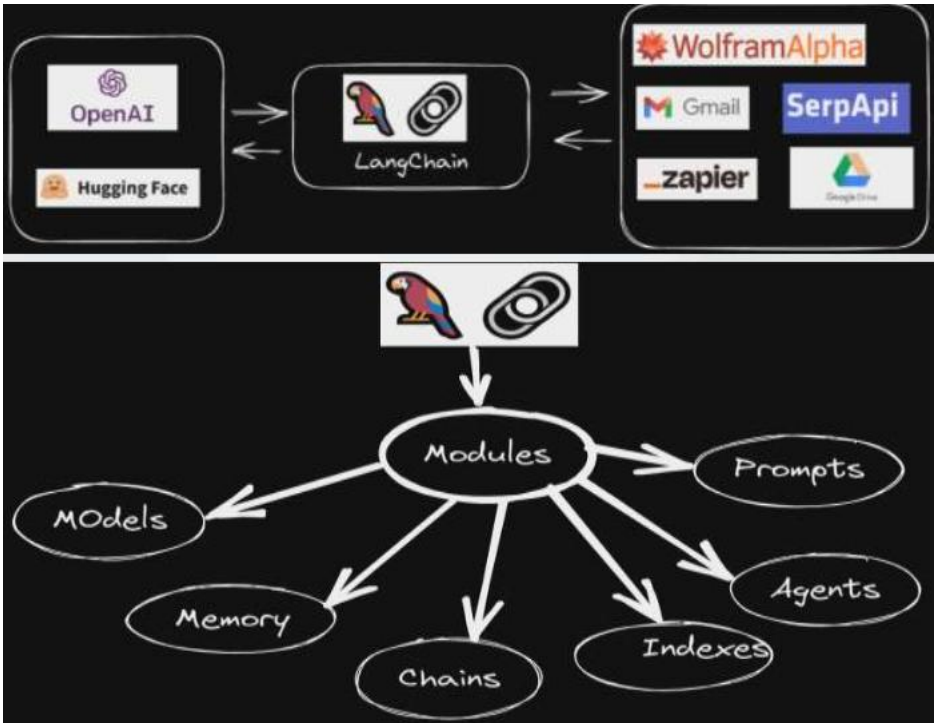


Table 2 Functions provided or integrations supported by Langchain.
(Source: Internal research report compiled by Qifu Capital)

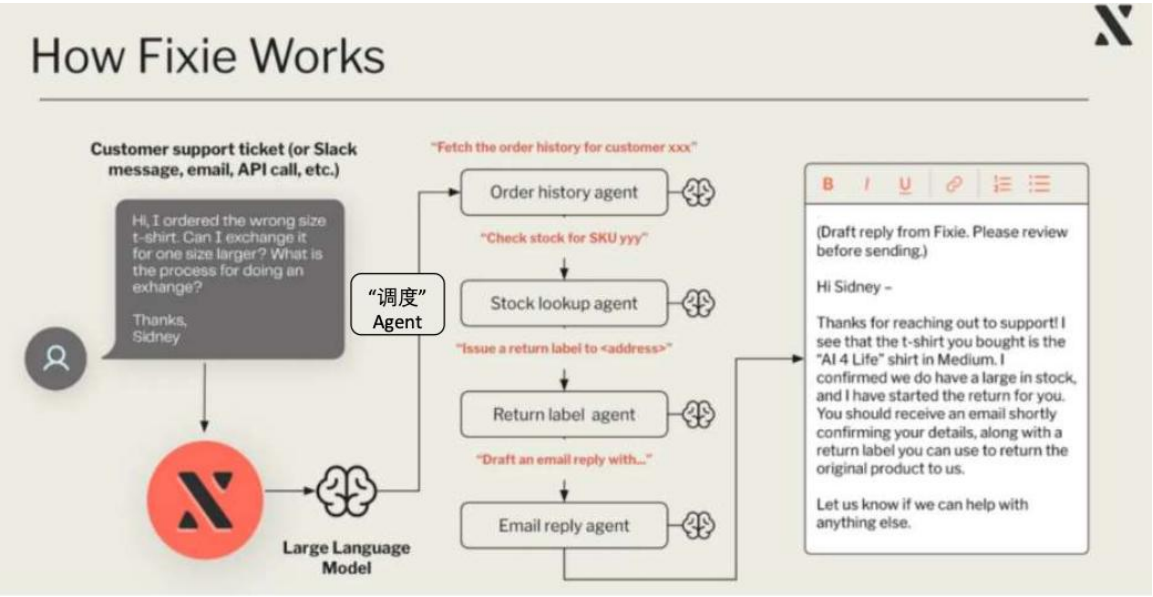
Dust aims to create a low-code platform for mainstream LLM application scenarios. Its core objective is to simplify the creation and deployment of large model applications, encompassing functions such as multi-model selection, external data source integration, version control management, parallel execution of

Features or supported integration provided by LangChain	
Data pre-processing	UnstructuredIO、Airbyte
Data index	GPT-Index
Doc & Text Splitter	Generic Recursive Text Splitter、Markdown Splitter、Python Code Splitter
Directional database and detection	FAISS、Pinecone、Weaviate、Elastic
Picture database	Chrome
External knowledge or operation	SerpApi、Searx、Wikipedia API、Wolfram Alpha
LLM API	OpenAI、Hugging Face Cohere、PaLM、GooseAI
Embedding engine	OpenAI、Hugging face、Cohere
Objectivity	Helicone、Prompt Layer、Weight&Biases
Application deployment	Streamlit、Huggingface、Streamship、Kookaburra
Model assessment data set	Hugging Face (truthful qa) 、Langchain Datasets
Model recovery structured and verified	Kor、Guardrails

multiple chains and examples, and template-based control execution. In its block-based modular architecture, common operations are encapsulated into blocks, each with fixed execution logic. Users only need to input parameters to get the results. By dragging and dropping blocks, users can build a simple LLM application.

Fixie focuses on automating workflows by integrating large models with external data sources and API tools, primarily targeting enterprise users. Fixie deploys models, external database connections, and API tools directly in the cloud. It breaks down tasks into multiple agents, with a scheduling agent understanding natural language and coordinating other agents to complete tasks across different systems. Fixie's core steps and few-shot learning approach include:

- Submitting user requests to the large model for task analysis and decomposition.
- Based on the model's output, the scheduling agent decides which agents to call.
- The scheduling agent calls agents responsible for querying order history and inventory data.
- Returning these data to the large model for analysis and determining if an exchange is supported.
- If an exchange is supported, the scheduling agent calls an agent to generate a return label.
- Simultaneously, another agent generates and sends an email to the customer.

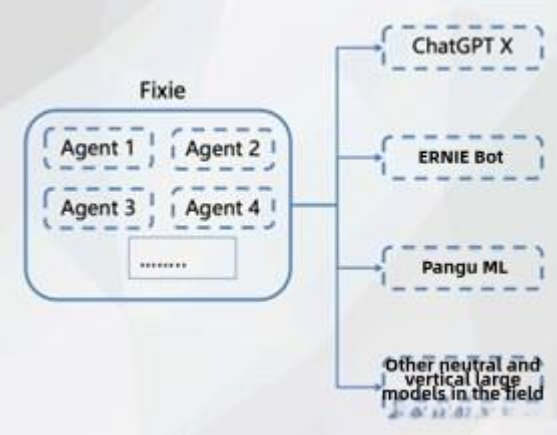


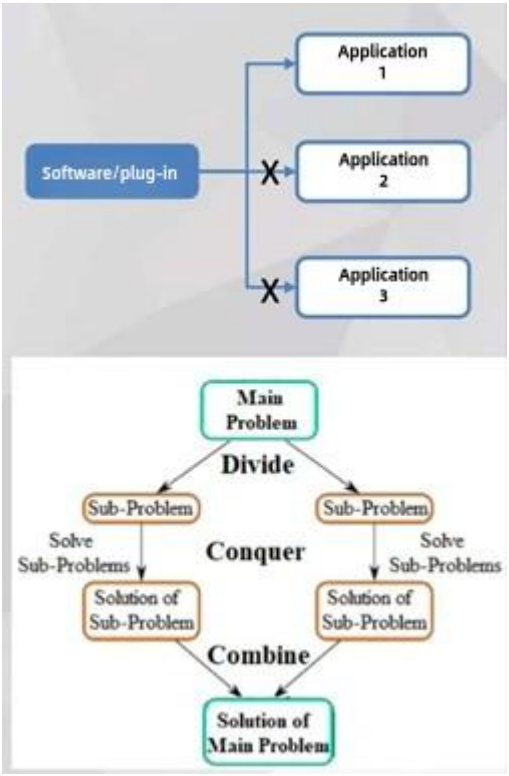
Fixie's advantages include the freedom for agents developed by independent authors to choose their underlying models, avoiding reliance on a single large model (as illustrated in Figure 22). Fixie's early-mover advantage could help it accumulate more third-party agents and enhance its agent scheduling capabilities, strengthening its position in the ecosystem and reducing over-reliance on ChatGPT. Traditional software or plugin development involves the software or plugin being shipped as a whole and applied to specific scenarios, making development and debugging more challenging, and leading to lower reuse rates. In contrast, under the agent concept, software and plugin development is divided into multiple modules, each of which can be developed independently and then combined, making development and debugging simpler, increasing modular reuse, and allowing agents to be reused in various applications. This aligns with the "Divide and Conquer" development philosophy and could reshape the software/plugin development ecosystem (comparison of traditional software/plugin development and agent-based

development models shown in Figure 23).

The evolution of AI paradigms has gone through two significant stages. The first stage involved the simple use and encapsulation of prompts. The second stage integrated external data sources to create applications like ChatGPT for X, such as ChatPDF and Chatbase. Currently, AI paradigms are entering a third stage, where AI is beginning to gradually replace human work. The core of this stage is to add API calls to external tools and self-iteration capabilities based on external data. Agents can be customized according to different work contents, and different agents can be coordinated through natural language. With the enhancement of large model capabilities, agents' abilities will also gradually improve. Combining general large models with specialized models to solve complex problems is likely to be a trend.

Currently, the ecological relationship between the middle layer and the foundational large models is not yet determined. The emergence of agents may reshape the AI and software development ecosystem. If the trend of multiple general large models combined with specialized models continues, the middle layer may strengthen its position in the development ecosystem through agents, changing the current dominance of OpenAI.



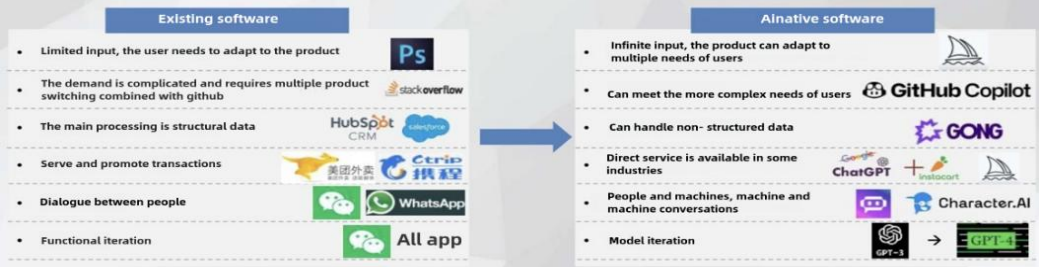


Appendix 9: Application Layer Analysis

In the application layer, the AIGC industry will undergo four levels of restructuring:

- **Interaction Restructuring:** Shifting from users adapting to products to products adapting to users, better meeting diverse and customized user needs.
- **Data and Information Restructuring:** Transitioning from structured data to unstructured data.
- **Service Restructuring:** In some industries, moving from organization and matchmaking to directly providing services.
- **Feedback Mechanism Restructuring:** Evolving from functional iteration to model iteration, allowing users to educate and correct AI, achieving self-learning through multiple attempts.

Under these restructuring influences, AIGC software will transform into AI-native software, with performance characteristics before and after the transformation illustrated in Figure 24.



The specific breakdown of AIGC applications in different industries is shown in Table 3.

Table 3 AIGC industry application analysis
(Source: Internal research report compiled by Qifu Capital)

Industry		Text generation	Audio Generation	Image Generation	Video/3D Generation
Software and information and technology services	B-end	Mainly generate general office process text	General voice broadcast and other functions	General purpose, marketing image generation/editing	General video generation
	C-end	/			
Communication	B-end	Mainly generate general office process text	General voice broadcast and other functions	Product design drawing generation/design aid tool	Simulation video
	C-end	/			
Ship	B-end	Mainly generate general office process text	General voice broadcast and other functions	Product design drawing generation/design aid tool	Simulation video
	C-end	/	/	/	/
Social contact	B-end	Instant match chat app	/	/	/
	C-end	Metaverse/virtual scene social			
Fitness	B-end	Marketing text generation	Marketing audio generation	Advertising/promotional graphics generation	Marketing advertising video generation, virtual fitness coach
	C-end	Personalized fitness plan	/	/	/
City government	B-end	Optimization of urban planning plans			
	C-end				
Research	B-end	Text optimization and extraction, scientific research idea assistance			
	C-end				

Industry		Text generation	Audio Generation	Image Generation	Video/3D Generation
Travel	B-end	Interactive tour guide, tour guide word generation	Tour guide explains speech generation	Advertising/promotional graphics generation	virtual tour
	C-end	Travel guide generation			
Food	B-end	Marketing text generation	Marketing audio generation	Advertising pictures/promotional pictures/packaging generation	Marketing advertising video generation
	C-end	Recommendations on what to eat (generate corresponding suggestions for the general direction)			
Logistics and transportation	B-end	Mainly generate general office process text	General voice broadcast and other functions	General image generation/editing	General video generation
	C-end	/	/	/	/
Aviation equipment	B-end	Design	General voice broadcast and other functions	Product design drawing generation/design aid tool	Simulation video
	C-end	/	/	/	/
Aerospace	B-end	Design	General voice broadcast and other functions	Product design drawing generation/design aid tool	Simulation video
	C-end	/	/	/	/
Consumer Electronics	B-end	Marketing text generation, product design optimization assistance	Marketing audio generation	Advertising graphics/promotional graphics generation/product design graphics	Marketing advertising video generation
	C-end	Comparative evaluation and performance understanding	/	/	/
Physical education	B-end	Sports press release writing	Guide/host audio generation	Advertising/promotional graphics generation	Guide/host video generation
	C-end	/	/	/	/
Education	B-end	Personalized lesson plan writing, personalized staff training, intelligent course design, helping schools manage and optimize students' learning progress	Pronunciation teaching guidance	Personalized teaching copy and picture generation	Teaching video generation and virtual experiment generation
	C-end	Lesson plan writing	Pronunciation teaching guidance	Personalized teaching copy and picture generation	Teaching video generation
Military industry	B-end	Mainly generate general office process text	General voice broadcast and other functions	Product design drawing generation/design aid tool	Simulation video
	C-end	/	/	/	/
Electronic Component	B-end	Mainly generate general office process text	General voice broadcast and other functions	Product design drawing generation/design aid tool	Simulation video
	C-end	/	/	/	/

Industry		Text generation	Audio Generation	Image Generation	Video/3D Generation
Agriculture, forestry, animal husbandry and fisheries	B-end	Mainly generate general office process text	General voice broadcast and other functions	General purpose, marketing image generation/editing	General video generation
	C-end	/	/	/	/
Traditional energy (coal, oil, natural gas)	B-end	Mainly generate general office process text	General voice broadcast and other functions	General image generation/editing	General video generation
	C-end	/	/	/	/
Mineral	B-end	Mainly generate general office process text	General voice broadcast and other functions	Generation of professional pictures such as topographic and geological maps of exploration areas	Mine structure simulation
	C-end	/	/	/	/
New Energy	B-end	Mainly generate general office process text	General voice broadcast and other functions	General image generation/editing	General video generation
	C-end	/	/	/	/
Energy storage	B-end	Mainly generate general office process text	General voice broadcast and other functions	General image generation/editing	General video generation
	C-end	/	/	/	/
Chemical insudtry	B-end	Mainly generate general office process text	General voice broadcast and other functions	General image generation/editing	General video generation
	C-end	/	/	/	/
Machinery and Equipment	B-end	Mainly generate general office process text	General voice broadcast and other functions	Design drawing generation	General video generation
	C-end	/	/	/	/
Food and Drink	B-end	Marketing text generation, intelligent customer service, personalized marketing content	Voice cloning (advertising endorsement)	Marketing posters/models/packaging	Marketing advertising video generation
	C-end	Personalized diet plan	/	/	/
Textile and Apparel	B-end	Marketing text generation, personalized marketing content	Voice cloning (advertising endorsement)	Marketing posters/model pictures/clothing design/packaging	Advertising video generation/product virtual display/digital human
	C-end	Text version clothing matching plan	/	Clothing matching diagram generation/virtual fitting	Clothing matching digital people
Home appliances	B-end	Marketing text generation, personalized marketing content	Voice cloning (advertising endorsement)	Marketing poster/model drawing/design drawing/packaging generation	Marketing advertising video generation
	C-end	Evaluation report	/	Personalized recommendations	/

Industry		Text generation	Audio Generation	Image Generation	Video/3D Generation
Car	B-end	Mainly traditional AI opportunities	Voice cloning (advertising endorsement)	Marketing poster/design drawing generation	Marketing advertising videos, simulations
	C-end	Evaluation report, itinerary optimization	/	/	/
Construction and Real Estate	B-end	Marketing text generation, architectural design plans, customer service chat	Marketing audio generation	Architectural/home decoration design drawing generation	House viewing video/videoization of decoration design
	C-end	Decoration and design plan	/	Architectural/home decoration design drawing generation	House viewing video/videoization of decoration design
Finance	B-end	Investment research reports, due diligence report generation, risk assessment and management, automated customer investment strategy generation	Roadshow, due diligence voice material generation	Trend chart/bar chart generation	AI Roadshow
	C-end	Investment research report (but involves licensing issues)	/	/	/
Insurance	B-end	marketing text writing, personalized product recommendations, after-sales service, claims processing	Marketing audio	/	/
	C-end	Insurance analysis reports, public opinion monitoring	/	/	/
Media	B-end	Media press release writing, personalized content recommendation	Virtual host/virtual anchor	Advertising image generation	Marketing advertising video generation, high-quality advertising generation,
	C-end	Personalized key content extraction	/	/	/
Entertainment (except games)	B-end	Automatic writing of novels, virtual stars, second creation	Music/song/virtual idol/dubbing/audio novel	Creative image generation	Video/animation/3D generation
	C-end	Automatic writing of novels	Music/song/virtual idol/dubbing/audio novel	Creative image generation	Video/animation/3D generation
Game	B-end	Game NPC plot generation and game scene design	game dubbing	Game original painting generation	3D game screen/video generation/secondary creation
	C-end	/	/	/	/
Medical Health	B-end	prescription generation, psychological consultation, drug analysis, and drug production optimization	Psychological Counseling (Voice)	/	Psychological consultation (digital human)/AI medical treatment
	C-end	Chatbot, simple prescription generation	/	/	/
Law	B-end	Legal consulting plan writing and contract writing	AI Defense Lawyer (Audio)	Universal image generation/editing	AI Defense Lawyer (Video/3D)
	C-end	Simple legal advice	/	/	/
E-commerce retail	B-end	marketing text generation, live broadcast skills, intelligent customer service/personalized recommendations,	AI selling	Marketing poster/model drawing/design drawing/packaging generation	Marketing advertising video generation/AI sales
	C-end	product comparison	/	/	/

AIGC will significantly contribute in three areas: increasing information capacity, serving industrial design, and acting as a personal assistant (providing suggestions for food and entertainment).

In **increasing information capacity**, AI software tailored to specific needs can be developed across various industries and sectors to enhance learning and work efficiency. Examples include:

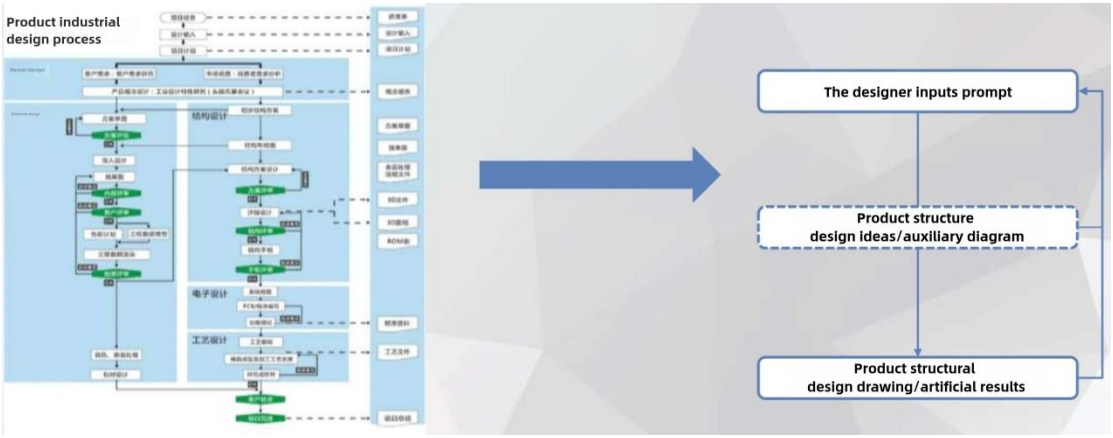
- **Higher Information Capacity Videos through Multimodal Generation:** Students can present questions to AI through text, voice, or images. AI will gather data based on multimodal inputs (text, voice, images, and videos) and generate higher-quality explanatory videos with higher information density per unit time.
- **Interactive Video Presentations:** During the explanation process, students can request clarifications ("I understand this part, but not this..."; "Please repeat this part, I didn't understand..."), achieving video interactivity.

- **Customized Suggestions and Plans through Multimodal Data Analysis:** Based on students' questions and communication, AI can analyse their other data and generate customized learning plans and suggestions.

In **industrial design**, current issues include cumbersome traditional design processes that mainly involve human participation, requiring multiple iterations by CAD designers, CAE simulation experts, and CAM process engineers. To address these issues, AIGC proposes the following business model:

- Extracting relevant data from industry/company-specific design and simulation databases and using models to learn and train on different features.
- Designers input design requirements as prompts, and the system automatically generates product 3D design or auxiliary diagrams.
- Based on historical simulation data, the system automatically generates simulation data for different scenarios.

This business model can be applied to various vertical segments such as mechanical equipment, automotive, aerospace, and defence industries.



In **lifestyle and entertainment**, the current issue is that consumers sometimes do not clearly understand their needs, having only a vague concept. AI can generate personalized activity suggestions based on consumers' potential desires and historical data (output can be multimodal). The idea can be compared to cocktail mixing: consumers describe a general taste they want (e.g., sour + refreshing + female-oriented). This also introduces new business models:

- Traditional companies like Dianping can build and train models based on existing user consumption/browsing preferences.
- Startups can build their own apps as traffic entry points or obtain data from other data providers.
- Users can input a general idea of what they want to do, and the app generates corresponding suggestions.

- The fee model can involve charging both users and merchants or only merchants.