

# Discussion on the Contradictions and Feasible Paths Between the Publishing Enterprise Model and the Basic Big Model in the View of Data Assets

Qiao Fangning<sup>1</sup>, Liu Feng<sup>2</sup>

<sup>1</sup>School of Journalism and Communication, Shanghai University, Shanghai, People's Republic of China

<sup>2</sup>(Corresponding author) School of Journalism and Communication, Shanghai University, Shanghai, People's Republic of China

Email address: panda197@163.com

**Abstract**—We are currently facing an increasingly complete state of mediatized society, in which every element cannot avoid being deeply affected by virtual networks and media, and even undergoes complete morphological transformation. The same is true for the publishing industry. As AI technology matures, one of the propositions that publishing companies have to face regardless of their scale is the impact from big models. In November, OpenAI released ChatGPT based on GPT-3.5, which instantly triggered the craze for big models at home and abroad. According to the "China AI Big Model Map Research Report" released at the end of May 23, 79 big models with parameters above 1 billion scale have been released nationwide. Therefore, this paper firstly details the contradictions and controversial points between the current enterprise big models and basic big models, analyzes them in terms of their logic and form, and builds a complete background framework. After that, the specific three parts of the content; the first part of the expansion explains the advantages and disadvantages of the basic model and enterprise model, the basic big model will be replaced by the prevalent rhetoric, but he still has the industry's small model cannot have the advantage. The basic model of higher accuracy, better performance and more stable, at the same time, it cannot adapt to the publishing industry's customized personalized needs, which also makes it become the basis of the big model of the disadvantage. Although the publishing enterprise's big model can meet the publishing enterprise's customized and personalized needs, but it does not have the enterprise volume of technical resources and talent reserves, only rely on the hands of the grasping the database resources cannot build the operation of the scientific big model. The second part of the story of the contradictions between the two, the basic model promotion and the development of enterprise big model there is a game in the contradictions highlighted on the basis of the final description of the publishing enterprise how to break through the difficulties, for the feasibility of the path of scientific exploration.

**Keywords**— Data assets, foundational large models, publishing enterprises, deep learning models, artificial intelligence.

## I. INTRODUCTION

Data asset refers to the data resources owned or controlled by individuals or enterprises, which can bring future economic benefits to enterprises and recorded in physical or electronic way. All industries under the data asset perspective have undergone transmutation in many aspects, and in publishing enterprises, data asset visualization is manifested in the construction of enterprise independent big model, fully utilizing the enterprise's own database information and

external arithmetic technology to enhance its own digitization level.

In recent years, publishing enterprises have begun to utilize the big model technology to improve their business ability and efficiency.

With the rapid development of information technology, the publishing industry has undergone great changes in terms of publishing carriers and publishing methods. For publishing organizations, publications are no longer a single resource, containing not only a collection of text, pictures, audio, video, CD-ROMs, attachments, metadata and other resources, but also related resources in the publishing process, such as structural information, typesetting information, design information, font information and so on. How to model all these resources and their relationships to realize the unified representation and reuse of these publishing resources is the problem facing digital publishing.

In order to better understand readers' needs, provide personalized recommendations and content editing, the publishing enterprise intends to build an intelligent recommendation system based on a large model. First, the enterprise collected a large amount of readers' data, including reading habits, preferences, historical behaviors and so on. Then, they used this data to train a deep learning model with the Transformer architecture for feature extraction and recommendation based on the self-attention mechanism. In the process of training the model, they used powerful computing resources and distributed training techniques to accelerate the training and optimization process of the model. At the same time, they also performed multiple iterations and tuning of the model to improve the accuracy and effectiveness of the recommendations.

Finally, the enterprise applied the trained big model to the actual recommendation system. By introducing this big model, they were able to better understand readers' interests and needs, generate personalized recommended content, and provide readers with a better reading experience. By building this intelligent recommendation system based on the big model, the publishing

## II. HOW TO CHOOSE BETWEEN "BASIC MODEL" AND SELF-BUILT ENTERPRISE MODEL

In the era of big data, the penetration of the whole publishing industry has become more and more profound. The competition and cooperation between publishing companies continue to contact new data application fields, and data assets have become the internal driving force for the transformation and development of the industry. The publishing companies that first use or even build the large model of independent enterprises will occupy the development opportunities. Can improve their competitiveness in the surging market waves will not be eliminated, and even have the opportunity to take the opportunity to win the top, standing in the industry development peak, to achieve self-enterprise benefits and corporate future planning reversal. Data assets have many applications and uses in the publishing industry.

The basic model refers to a large-scale language model based on deep learning technology, which can automatically learn and process a large amount of natural language text data, and can generate natural and smooth language text. Since 2023, ChatGPT has led a new round of global large model innovation, China's "2023-2024 China Artificial Intelligence computing Power Development Assessment Report" shows that as of October 2023, China has released more than 200 large models (including basic large models and industry large models), has entered the "100 models exhibition" era.

While talk is rife that the underlying big model will be replaced, it still has advantages that smaller models in the industry cannot.

First of all, after the basic model exceeds a certain parameter threshold, the AI effect will no longer be a random probability event, and its accuracy rate will be higher and higher, and the degree of intelligence emergence will be higher and higher. Secondly, the basic large model improves its applicability in learning from massive and multi-type scenario-based data and provides possibilities for subdividing scenarios. Finally, compared with small-style production with high average cost, long cycle and low utilization rate of small-model R&D, large-scale large-model adopts factory-style production, reducing dependence on specific data, significantly reducing labor costs and improving R&D efficiency.

At the same time, the basic large model also has some disadvantages, the most prominent is that the basic model still does not "understand" many specific needs of the publishing industry, can not adapt to personalized production scenarios. This is determined by its homogenous characteristics, the ability of the base model is the center and core of intelligence, and therefore does not have the attributes of the industry. However, in general, the basic large model is the basis of the industry model, and consolidating the foundation of the basic model is very helpful for the training of diversified and high-quality industry models.

Second, data assets have become an intrinsic driver of industry transformation and development. In the era of big data, the penetration of the whole publishing industry has become more and more profound. The competition and

cooperation between publishing companies continue to contact new data application fields, and data assets have become the internal driving force for the transformation and development of the industry. The publishing companies that first use or even build the large model of independent enterprises will occupy the development opportunities. Can improve their competitiveness in the surging market waves will not be eliminated, and even have the opportunity to take the opportunity to win the top, standing in the industry development peak, to achieve self-enterprise benefits and corporate future planning reversal. Data assets have many applications and uses in the publishing industry.

Third, the development of the industry has a realistic need for the enterprise large model

On the one hand, the whole process of online reading such as electronic books, KINDLE, wechat reading is becoming more and more abundant, Douban reading, Tomato reading, etc., in addition, lazy people listening to books, Himalayas and other apps that can listen to books are also emerging in endlessly. Corresponding to this, paper books are heavy and not easy to carry, inconvenient to store, and have the characteristics of single reading. Compared with this, more and more readers prefer to choose electronic reading. As can be seen from the chart of the per capita reading number of Chinese people from 2010 to 2022 by the Chinese Academy of Press and Publication, the number of online readers in China is constantly increasing, and its growth rate is climbing very fast. In addition, with the popularity of short video and fragmented reading era, publishing institutions are faced with the loss of a large number of consumers, and they need to make full use of their hard-won high depth data through digital and data forms, innovate book publishing forms, activate the due value of books, and ultimately help publishing enterprises to revitalize revenue and "revive" with blood. On the other hand, readers in the new era still have a sense of awe and respect for paper books, and there are still a large number of readers who choose to support paper books and physical bookstores. Finally, the booming development of online literature is also closely related to a large number of online authors. The report of Zhiyan Consulting Research Institute shows that the number of digital reading content creators will reach 9.29 million in 2022, and the scale of digital reading authors will continue to expand. It can be seen that building a large enterprise model is not only the needs of the publishing industry to seek its own development, but also the call of a large number of readers, and constantly provide greater growth space for content creators.

## III. THE BASIC CONDITIONS FOR THE DEVELOPMENT OF LARGE MODEL OF PUBLISHING ENTERPRISE

The level and intensity of the large model that can be built by the resources at the technical level are also different. Publishing enterprises use artificial intelligence technology and artificial intelligence new media in semantic analysis, data analysis, autonomous machine learning and other aspects. At present, the basic idea of semantic publishing modeling is to structure and semanticize the publication content, which mainly covers the content elements, structural elements and

the whole publishing process. The model consists of six parts: document constitution, embedded part, reference and reference, executable and generative. The object and relation are described and serialized by RDF for each part.

Computing power is the ability of computer equipment or computing/data centers to process information, and is the ability of computer hardware and software to work together to perform certain computing needs. Computing power is characterized by intensification, integration, synergy, value, etc. Through numerical integration, the transmission, processing, utilization and activation of data resources, data products and model algorithms can be effectively operated, which can maximize the operational efficiency of the publishing industry and accelerate the digital transformation of the internal operation of publishing enterprises.

If you want to construct a large model, you must first build a large scale GPU computing power cluster, usually 100 billion parameter scale can be called a large model. For example, Wenxin large model (ERNIE 3.0 Titan) parameters up to 260 billion. This is difficult for ordinary publishing companies to achieve, and the bandwidth, video memory, the number of servers, etc. of publishing companies are not enough to support such a huge computing power technology.

In addition to "big data", the vertical and professional data accumulated by the publishing industry over the years plays a core role in the construction of professional large models in the publishing field.

It can be seen that the self-built digital resource model of publishing enterprises mainly focuses on the collection, statistics and analysis of publishing internal information, and uses large language models such as chatgpt for dialogue generation, text summary generation, question and answer system, etc. It is mainly used in the field of publishing cooperation, topic selection planning, manuscript review and proofreading, digital human corpus supply, customer service and sales, institutional management and decision-making. On the back end of data realization, transformation and marketing, there is insufficient experience and less application. Moreover, this kind of large language model has obvious shortcomings in logical reasoning, cultural correlation, meaning mining, etc., so it is easy to cause text ambiguity and intellectual property disputes

#### IV. CONTRADICTIONS IN THE CONSTRUCTION OF LARGE MODELS FOR PUBLISHING ENTERPRISES IN THE CONTEXT OF DATA ASSETS

"The conditions required for large enterprises to possess foundational models."

Data assets are the result of the growth of big data and represent a phased product of continuous evolution in the big data realm. Within publishing enterprises, industry data serves as a crucial resource for model construction. The long-standing accumulation of industry behavioral data, regularity data, user behavior data, as well as data from authors themselves, all constitute unique advantages for publishing enterprises in building large models. These resources not only enhance resource utilization but also facilitate profit generation and revitalize old inventory.

Despite the vast amount of data possessed by publishing enterprises, their weaknesses in technical expertise and computing power are akin to the fatal flaw of a barrel with a short plank. According to the "2023 National Reading Trend White Paper," lack of time, energy, and difficulty finding content of interest are the top three obstacles to cultivating reading habits. Presently, publishing enterprises are unable to address these user dilemmas effectively, leading to user attrition.

In this context, the construction of large models by publishing enterprises is crucial, yet impeded by technical and computing power issues. For foundational model enterprises like Baidu and iFlytek, they possess powerful computing and technical capabilities but lack certain databases. There exists a certain conflict of interest between the two parties, making it difficult to reach a consensus. Essentially, the conflict and contradiction lie between building foundational large models and enterprise-specific small models.

Domestic basic large model enterprises led by Baidu are able to construct foundational large models that rival the GPT system, and their substantial technical, equipment, and talent conditions are inseparable from this capability.

Firstly, they possess strong technological capabilities, robust computing management, and maintenance abilities, ensuring relatively stable data security. The efficient operation of computing power data centers relies heavily on substantial electrical support, necessitating a large number of electrical grids, motors, and other equipment. Enterprises with foundational models not only have sound intelligent infrastructure but also strong proprietary capabilities, technological innovation, and creativity, making their overall computing power stronger compared to publishing enterprises.

Publishing enterprises started digitization relatively late and progressed slowly. Their competition with large internet enterprises is relatively minor, with lower technological requirements and a lower demand for updates and upgrades. Therefore, it is more challenging for publishing enterprises to independently build foundational models, while it appears relatively easier for large enterprises like Baidu and iFlytek to do so.

Baidu Intelligent Cloud's Qianfan Large Model Platform is a one-stop enterprise-level large model platform that offers advanced generative AI production and applications, providing a full development toolchain. The Qianfan AI native application workspace contains general large models, industry-specific large models, Baidu Wenxin large models, and various third-party large models, offering a wide range of application scenarios and paradigms with built-in high-quality large model capabilities. These are capabilities that publishing enterprises find difficult to match.

In 2023, China's large model field saw continuous positive progress. Recently, the results of the first official "Large Model Standard Compliance Test" were announced nationwide, with four domestic large models—Baidu Wenxin Yiyuan, Tencent Hunyuan, 360 Zhinao, and Alibaba Cloud Tongyi Qianwen—passing the initial test. This signifies that these companies' relevant technologies meet national standards in terms of requirements, universality, and

intelligence, allowing them to be applied in a broader range of scenarios in the future to meet the needs of various industries and users.

Secondly, although large enterprises have mature technological conditions and rapid updates, having just the "dumpling wrapper" is not enough to make dumplings; the industry data of publishing enterprises, the essential "dumpling filling," is indispensable. Lack of industry-specific foundational data makes it challenging for large enterprises, represented by internet companies providing technological and talent support through basic models, to coordinate intellectual property rights regarding knowledge and user data with publishing enterprises. Thus, there is still no relatively fair and reliable intermediary platform that can provide industry data for large enterprises. Taking Baidu Wenxin Yiyuan as an example, training large models requires three layers: the chip layer, framework layer, and model layer. However, the lack of industry-specific foundational data in the model layer means that large enterprises' basic models only have the "dumpling wrapper" without the "dumpling filling." This implies the inability to obtain industry large models with complete data and profiles, thus unable to construct complete large models that can provide comprehensive services to users. Large models also encounter gaps and errors during the process of updates, iterations, and self-learning due to insufficient sample quantities, leading to wastage of computing power and hardware cluster resources.

Taking ChatGPT as an example, its self-updating and iteration also require diverse and high-quality training data. In terms of data diversity, GPT-1 mainly uses training corpora such as books, like BookCorpus. GPT-2 incorporates news data like Reddit links, with high-quality text standards and includes some social data from everyday communication. With GPT-3, the model's data scale increases by tens of times, with the addition of datasets like Reddit links, Common Crawl, WebText2, Wikipedia, significantly enhancing data diversity. In the GPT-4 stage, GitHub code, dialogue data, some mathematical application questions, and even multimodal data are introduced.

Thirdly, regarding data quality, taking the GPT-3 model as an example, 75% of the training corpus consists of English data, 3% is Chinese, with additional datasets in languages such as Spanish, French, and German. These learning corpora are sourced through various means including public data (such as Wikipedia, Baidu Baike, Weibo, Zhihu, etc.), open-source datasets, web scraping (GPT-3 was trained on approximately 31 billion web pages, comprising around 300 billion words), and private datasets (like OpenAI's WebText dataset, which collected 8 million highly upvoted articles from the Reddit platform, totaling about 15 billion words). Among these corpora, English data from public sources are more abundant and of higher quality.

Despite the high quality of the aforementioned data, which includes data from sources like Wikipedia, books, and academic journals, it only accounts for 17.8% of the dataset. However, its weight in model training reaches 40%, highlighting the continuous pursuit of data quality and its significance in training.

In response, some industry analysts believe that when a certain critical threshold of high-quality data is reached, it will significantly narrow the accuracy gap between different algorithms, ultimately determining the quality of model training. This not only makes training more efficient but also substantially reduces training costs.

In contrast, there is a scarcity of open-source high-quality Chinese data, especially in categories such as encyclopedic articles, question answering, literature, academic papers, newspapers, and magazines, which are essential for building foundational large models. Additionally, professional data services in China are still in their early stages, and high-quality datasets processed, cleaned, and annotated for AI model training are relatively scarce.

Therefore, the lack of high-quality and diverse training data has become one of the core pain points and the most significant challenge for domestic foundational model training. Addressing this challenge requires obtaining a large amount of open-source data from the industry. Directly acquiring cleaned and more professionally curated data from publishing enterprises would greatly benefit the improvement of large model quality, making them more comprehensive.

#### V. THE PROMOTION OF FOUNDATIONAL MODELS AND THE DEVELOPMENT OF ENTERPRISE-SCALE MODELS INVOLVE A GAME OF STRATEGY.

For the basic large model, the basic model can only be realized and economic value can be realized by promotion, so the basic model needs to find a way to promote. The promotion of itself requires the cooperation or even cooperation of small business models. Even with push removal, the process of updating it is extremely difficult in the absence of a database. In addition to the lack of high-quality data, it is also easier to generate conflicts and conflicts of interest with enterprises to build their own models. For enterprises, although the quality of the database for building their own small models is high enough and sufficient, because they are not specialized in technology development and intelligence, their own resources for building small models are not sufficient, and their own small models are not enough to protect themselves. This makes it difficult for the two to quickly find a running-in point in the market, so that it is difficult to reach a moderate cooperative relationship.

The main participants in this game can be divided into two types. Firstly, there are intelligent enterprises represented by companies like Baidu and iFlytek, which possess the capability to develop "foundational models" and strive to seize the initiative (to create China's own GPT). ChatGPT was released on November 30, 2022, as an AI-driven natural language processing tool. Within just five days of its launch, the number of registered users exceeded one million, and by the end of January 2023, the monthly active users had surpassed 100 million, making it the fastest-growing consumer application in history. Faced with such urgent competition, to prevent ChatGPT from monopolizing the market, leading Chinese internet companies such as Baidu and Alibaba have embarked on a deeper process of intelligence, continuously building foundational large models to enhance their

independent AI research and development capabilities. Examples include Baidu's Wenxin One Word large model and iFlytek's Xunfei Starfire Cognitive large model, representing the rapid progress of China's foundational large model construction.

The other type of participant is large publishing conglomerates. These entities possess years of accumulated reader data, literature, author information, and other data, which are characterized by comprehensive content, high authenticity, and a combination of historical and contemporary relevance. They provide extremely powerful assistance for artificial intelligence learning and updates. Additionally, large publishing conglomerates are facing the challenges of digitalization and adapting to new user behaviors, necessitating the construction of large models to address these issues.

These two types of participants engage in a game of competition, which ultimately influences the paths chosen for the development of large models.

Firstly, companies like Baidu and iFlytek tend to focus on developing "foundational models" and incorporate data and requirements from various industries such as publishing, film, and television into their development scope. Acquiring high-quality professional data accumulated from different industries holds significant value for them.

Secondly, publishing enterprises face different choices in the development of large models. One option is to follow the path of "foundational models" by integrating their publishing data into the overall data pool, thereby enhancing the intelligence level of foundational models in publishing scenarios. Another option is to build their own enterprise-scale models, which, although applicable to a narrower scope, can meet the basic needs of publishing practices.

The core of the above two paths for publishing enterprises lies in whether to "cede" their data to foundational models or to keep their publishing data within the boundaries of their own business scope. Thus, the choice between "foundational models" and building enterprise-scale models, as well as the central issue in the game between the two types of participants, revolves around data assets.

## VI. EXPLORING FEASIBLE PATHS FOR BUILDING LARGE MODELS IN PUBLISHING ENTERPRISES.

Scenario 1: Collaboration between Foundational Large Models and Publishing Enterprises. Publishing enterprises "cede" data. Publishing enterprises cannot remain isolated with their own databases, as only through continuous openness can they better utilize their data assets and realize their value. However, this ceding of data does not mean relinquishing complete ownership rights; rather, it involves transferring resources while safeguarding one's own rights.

Publishing enterprises should reach agreements with foundational model providers on various aspects such as the authority to cede data resources, defining data resources, channels for data acquisition, usage methods and modes, data distribution, and customization. This ensures that their rights are as protected as possible while achieving a "win-win" situation.

When collaborating with companies that possess foundational models, it is essential to establish protective mechanisms in advance, consult with law enforcement agencies, and utilize legal means to defend data rights. Additionally, dedicated departments should be established within the corporate management structure to monitor data throughout the collaboration process, preventing situations where data loss occurs without awareness.

Scenario 2: Embarking on the Path to Constructing Enterprise Large Models. If publishing enterprises wish to address their deficiencies in technology and computational power, it is better to teach them to fish than to give them fish. Rather than relying on external aid, publishing enterprises can, to a certain extent, bridge the gap between themselves and the digital publishing era by constructing their own large models. Therefore, addressing the shortcomings in technology and computational power is a crucial step. Only with the "dough wrapper" can one effectively utilize their "dumpling filling".

Firstly, attracting or recruiting a large number of intelligent technical talents is essential. Only with a rich pool of comprehensive talent can cross-industry and cross-perspective learning and work be supported. Secondly, utilizing existing paid large models for simple model construction or learning is viable. Currently, there are many large model products available on the market, some of which can even be customized for data analysis, marketing, and other operations specific to the enterprise through payment. Thirdly, fundamentally reforming the organizational mechanisms of publishing enterprises is necessary, making thorough changes in development ideology, perspectives, and concepts, targeting internal pain points, and reforming the existing institutional mechanisms of publishing enterprises to ones that are more adaptable to digital construction. This transformation aims for flexibility, agility, intelligence, and interactivity. Shifting from being author-centric to reader-centric, from technology being an embellishment to being the primary developmental support, and from emphasizing monotypic talent to valuing multifaceted talent.

Constructing enterprise large models should not merely be seen as technical and computational challenges; instead, it should involve comprehensive "innovation" and restructuring of the publishing industry.

Scenario 3: Collaborative Efforts between Publishing Enterprises and Companies with Foundational Large Models. The relationship between publishing enterprises and companies possessing foundational large models is not one of opposition, but rather a collaborative and evolving parallel relationship. They can establish cooperative relationships, clarifying the boundaries of database usage through continuous experimentation, and balancing the interests and rights of both parties. Publishing enterprises can leverage the computational power, technology, and talent advantages of the other party to refine and filter their own databases, while companies with foundational models can gain certain usage rights to vast databases, enabling rapid evolution of their foundational large models.

## VII. CONCLUSION

The digital publishing process has been underway in foreign countries for a longer time, with research progressing relatively rapidly. In China, only a few academic database service providers such as CNKI (China National Knowledge Infrastructure) and Wanfang Data have embarked on the construction of large models. There are relatively few publishing enterprises that have constructed independent large models or collaborated with foundational large models. Furthermore, there is a lack of related practical research, and the digitalization process of many publishing enterprises, especially small and medium-sized ones, remains slow. Despite the maturity and relatively low cost of many data analysis technologies, difficulties in implementation and collaboration persist.

Silicon Valley's tendency to merge innovative small companies and monopolize markets seems like an inevitable historical trend. However, Marxist theory suggests four stages of capitalist development: from primitive capital accumulation to monopolistic imperialist economies, and from mixed economies to neoliberalism. Although foundational large models currently exhibit dominance, historical patterns indicate that new forms of cooperation representing neoliberalism will emerge.

"Divided we spark, united we blaze." The friction between Chinese publishing enterprises and large model companies can not only enhance their individual strengths but also collectively drive the construction of models and the level of autonomous intelligence in China.

## REFERENCES

- [1] Geying, Y., & Shen, C. (2019). Research on the Evaluation System of Competitiveness of Digital Publishing Industry: A Case Study of Jiangsu Province. *Market Weekly*, (11), 49-51
- [2] Che, W., Dou, Z., Feng, Y., et al. (2023). Natural Language Processing in the Era of Large Models: Challenges, Opportunities, and Development. *Chinese Science: Information Science*, 53(09), 1645-1687.
- [3] Liu, T. (2023). Discussing Large Language Models and Their Applications from ChatGPT. *LanguageStrategicResearch*, 8(05), 14-18. DOI: 10.19689/j.cnki.cn10-1361/h.20230501.
- [4] Yu, J., Liu, Y., & Wu, C. (2023). The Transformation and Governance of Artificial Intelligence Large Models. *Chinese Public Administration*, 39(04), 6-13. DOI: 10.19735/j.issn.1006-0863.2023.04.01.
- [5] Lei, X., & Shao, B. (2023). Research on Digital Literacy Education Embedded in Artificial Intelligence Content Generation under Large Models. *Modern Information*, 43(06), 99-107.
- [6] Li, G., Wang, Z., He, X., et al. (2023). From ChatGPT to Multimodal Large Models: Current Status and Future Prospects. *China Science Foundation*, 37(05), 724-734. DOI: 10.16262/j.cnki.1000-8217.20231026.001.
- [7] Zhao, X., Dou, Z., Wen, J. (2023). Research Trends in Information Retrieval in the Era of Large Language Models. *China Science Foundation*, 37(05), 786-792. DOI: 10.16262/j.cnki.1000-8217.20231026.007.
- [8] Chen, L., Zhang, S., Yu, K. (2023). Cross-Modal Language Large Models: Progress and Prospects. *China Science Foundation*, 37(05), 776-785. DOI: 10.16262/j.cnki.1000-8217.20231026
- [9] Xia, R., Li, P. J. (2023). Development and Application of ChatGPT Large Model Technology. *Data Acquisition and Processing*, 38(05), 1017-1034. DOI: 10.16337/j.1004-9037.2023.05.002.
- [10] Zhai, Y., Li, J. (2022). Reflection on the Development Path of AIGC: New Opportunities for Popularizing Large Model Tools. *Internet World*, (11), 22-27.