Overview of Research on Large Model-Driven AI Application: Technical Adaptation Paths, Scenario Innovation Paradigms, and Industrial Empowerment Value

Qingtao Wu, Deming li
School of Computer Science,Zhengzhou University of Aeronautics, Zhengzhou,China

Abstract: This paper takes the large model-driven AI landing application as the research object and conducts a systematic review around the three-dimensional main line of "Technical Adaptation - Scenario Innovation - Industrial Empowerment". It sorts out the core technical adaptation paths consisting of model lightweighting, deployment mode adaptation and the trustworthy AI system, analyzes the scenario innovation paradigms in the fields of government services, industrial operation, industry and people's livelihood, and expounds the empowering value of large models at the levels of enterprise cost reduction and efficiency improvement, industrial format upgrading and macro industrial synergy. Meanwhile, it points out the core challenges faced by the current landing of large models, such as technical compliance and scenario-industry adaptation. Finally, it prospects the technical development trends of lightweighting, verticalization and trustworthiness, as well as the application directions of in-depth development of segmented scenarios and industrial ecological synergy, providing a theoretical reference and practical framework for the industrialization landing of AI technology.

Keywords: Large Models; AI Landing; Technical Adaptation; Scenario Innovation; Industrial Empowerment

1. Introduction

In recent years, breakthroughs have been made in large model technology centered around the Transformer architecture, with parameter scales jumping from the tens of billions to the trillion level, continuously refreshing technical indicators in fields such as natural language processing and computer vision. As technology maturity increases, large models have moved from the laboratory to industrial implementation. The global market size of enterprise-level large models is expected to exceed $80 billion by 2025. Domestically, policies such as the "Opinions on Deepening the Implementation of the 'Artificial Intelligence +' Initiative" and the "Interim Measures for the Management of Generative Artificial Intelligence Services" have been introduced in a concentrated manner, explicitly positioning the application and implementation of large models as the core support for digital economy transformation.

Currently, large models face multiple practical pain points in the process of industrial application: at the technical level, the contradiction between high computing power consumption and high deployment costs of large models and the limited resources of small and medium-sized enterprises is prominent; at the scenario level, most applications remain at the general service level, with insufficient cases of deep adaptation to vertical industries, and a serious phenomenon of homogenization; at the industrial level, technology is disconnected from business processes, the empowerment effect is difficult to quantify, and the industrial chain collaboration mechanism is not yet sound. Against this backdrop, this review focuses on the three-dimensional main line of "technical adaptation - scenario innovation - industrial empowerment", systematically sorting out the core logic and practical path of large model implementation. It not only provides theoretical references for technological research and development in the computer field but also offers a practical framework for the digital transformation of traditional industries, which has important academic value and practical significance.

2. Core Technology Adaptation Path for Large Model-Driven AI Implementation

2.1 Model Lightweight Adaptation Technology

2.1.1 Model quantization technology

Model quantization technology reduces storage occupation and computational load by decreasing the numerical precision of model parameters and activation values. Its core is divided into three categories: symmetric quantization, asymmetric quantization, and mixed-precision quantization. Symmetric quantization is suitable for scenarios where parameter distributions are uniform, asymmetric quantization can adapt to non-uniformly distributed data through offset adjustments, and mixed-precision quantization combines the advantages of different precisions to achieve a balance between performance and efficiency. In cold wave warning scenarios, researchers optimized the prediction model based on CNN-LSTM using INT8 asymmetric quantization technology, converting the original FP32-precision model parameters to INT8 precision, while compensating for performance loss through quantization-aware training (QAT). After optimization, the model prediction speed increased from 3 seconds per prediction to 0.8 seconds per prediction, meeting the real-time warning requirements; hardware power consumption decreased from 17W to 2.9W, adapting to the low power consumption requirements of edge computing devices; the prediction accuracy remained at 92.3%, with only a 0.7% difference from the original model. This technology has been widely applied in scenarios with high real-time requirements such as meteorological warnings and intelligent transportation.

2.1.2 Adaptation of MoE and MHA architectures

The Mixture of Experts (MoE) model utilizes a dynamic scheduling mechanism of "gated network - expert sub-models" to activate only the sub-models relevant to the input, thereby reducing computational costs while maintaining model capacity. Its core advantage lies in decomposing general tasks into multiple specialized tasks, with each expert sub-model focusing on feature learning in a specific domain. In the context of multi-modal data processing in power grids, the 16-expert MoE model allocates tasks such as load forecasting, fault diagnosis, and anomaly detection to different sub-models. The gated network dynamically assigns weights through the Softmax function, ultimately achieving a 5.8% improvement in prediction accuracy, a 31% reduction in false alarm rate, and a 40% increase in computational efficiency.

The Multi-Head Attention (MHA) mechanism captures the semantic associations and structural features of input data from different dimensions through multiple parallel attention heads. Its core formula is $\mathrm{Attention}(Q,K,V)=\mathrm{softmax}(\frac{QK^T}{\sqrt{d_k}})V$, where d_k represents the dimension of the attention head. In the context of power equipment defect report analysis, MHA utilizes 12 attention heads to separately capture association information from dimensions such as equipment type, fault phenomenon, and operating environment. By combining positional encoding and layer normalization techniques, it reduces the time required for fault root cause localization from 30 minutes to 10 minutes, and increases the accuracy of root cause identification to 89%, significantly enhancing the technical adaptability in complex scenarios.

2.2 Deployment Mode Adaptation Strategy

2.2.1 Privatized deployment

Privatized deployment refers to the operation of large models on local enterprise servers, private clouds, or proprietary hardware, with data storage and computation completed internally within the enterprise, effectively avoiding data leakage and privacy breaches. Its technical architecture comprises four levels: hardware layer (servers, GPU clusters), system layer (operating system, virtualization software), platform layer (large model training and inference platform), and application layer (industry solutions). This model is primarily suitable for highly regulated industries such as finance, healthcare, and government affairs, and can meet the compliance requirements of the "Cybersecurity Level Protection Regulations" and other regulations at or above the third-level protection. Taking the intelligent risk control large model of a state-owned bank as an example, the privatized deployment solution utilizes Huawei Kunpeng servers and Ascend GPU clusters. It achieves physical separation of training data and business data through data isolation technology, ensures operational traceability through access control and audit logs, effectively prevents data leakage risks, and simultaneously meets the regulatory requirements of the China Banking and Insurance Regulatory Commission for financial

data security.
2.2.2 Hybrid deployment and computing power sharing
The hybrid deployment adopts a collaborative architecture of "cloud + local", deploying general computing tasks on public clouds and sensitive data processing and core business operations on local servers. Its core advantage lies in balancing data security and deployment costs. In the intelligent operation scenarios of cross-regional chain enterprises, the hybrid deployment solution deploys general tasks such as customer personas and marketing recommendations on Alibaba Cloud, while storing sensitive data such as financial data and customer privacy information on local servers. Secure communication between the cloud and the local is achieved through an API gateway, and data transmission employs AES-256 encryption technology to ensure data transmission security. This model can reduce enterprise hardware investment costs by more than 30%, while meeting data compliance requirements.
The computing power sharing model is based on blockchain, edge computing, and other technologies, building a distributed computing resource pool. Small and medium-sized enterprises (SMEs) can use large model computing power through on-demand leasing, effectively reducing the threshold for technology use. Its core mechanism includes three links: computing node access, resource scheduling, and billing and settlement. The automated execution of computing power transactions is achieved through smart contracts. Alibaba Cloud's "Feitian Smart Computing Platform" and Tencent Cloud's "Xingmai Computing Power Pool" have both established mature computing power sharing ecosystems. SMEs can use large model training and inference services through pay-per-use, monthly subscription packages, and other methods, reducing the cost of computing power usage by more than 50%, promoting the widespread adoption of large model technology.

2.3 Trustworthy AI Technology Adaptation System

Addressing issues such as hallucinations, insufficient interpretability, and privacy leakage in the implementation of large models, the industry has established a four-dimensional trusted AI technology system, namely "infrastructure layer - platform service layer - application layer - governance layer". The infrastructure layer, centered around security chips and trusted computing modules (TCMs), provides hardware-level security guarantees. The platform service layer integrates privacy computing technologies such as federated learning, differential privacy, and homomorphic encryption to achieve "usable but invisible" data. The application layer employs explainable AI (XAI) algorithms, such as LIME and SHAP, to enhance the transparency of model decision-making. The governance layer constructs compliance auditing, risk assessment, and emergency response mechanisms to ensure full process controllability.
Haiyun'an's trusted large model platform adopts federated learning technology to build a distributed training framework. It achieves joint training of data from multiple enterprises through horizontal federation and achieves multi-dimensional data complementarity through vertical federation, effectively preventing poisoning attacks on training data. NSFOCUS's large model security detection system can monitor risks such as hallucination content and sensitive information leakage in model outputs in real time. It achieves a risk warning accuracy rate of over 95% through dynamic threshold adjustment. Practice has shown that the trusted AI technology system can reduce the compliance risks of large model deployment by 60% and decrease the incidence of privacy leakage incidents by 75%, providing crucial support for the large-scale deployment of large models.

3. Scenario Innovation Paradigm Driven by Large Models for AI Implementation

3.1 Innovation in Intelligent Government Service Scenarios

3.1.1 Intelligent document and file management
In the government office scenario, large models leverage natural language generation (NLG), text classification, information extraction, and other technologies to achieve intelligent processing of official documents throughout the entire process. Density Technology's government large model solution is based on a BERT fine-tuned model, constructing a comprehensive system covering the entire process of official document generation, review, classification, and archiving. In the official

document generation phase, after users input key information such as the document theme and core requirements, the model can automatically call upon the government official document template library to generate a preliminary official document that adheres to format specifications and rigorous language. It supports one-click modification and format optimization, boosting the efficiency of official document generation by over 70%. In the document classification and archiving phase, the model automatically assigns multi-dimensional labels such as organization, theme, and time to files through keyword extraction, topic clustering, and other technologies, constructing an intelligent retrieval engine. The document retrieval time has been shortened from an average of 30 minutes to 2 minutes, with a retrieval accuracy rate of 93%. This solution has been implemented in a provincial government service center, processing over 100,000 official documents annually, boosting the efficiency of document archiving by 80%, and significantly reducing the workload of government personnel.

3.1.2 Public opinion monitoring and project approval

The large model has established a closed-loop system of "data collection - analysis and judgment - decision support" in government affairs and public opinion monitoring. Through web crawling technology, multi-source data from social media, news websites, government message boards, etc. are collected. After data cleaning and standardization processing, they are input into the large model for sentiment analysis, topic clustering, and trend prediction. The model can capture the public's attitude towards policy implementation and public events in real time, identify potential public opinion risk points, and generate visual reports including public opinion heat, sentiment distribution, and development trends, providing data support for government decision-making. After the launch of a prefecture-level government affairs and public opinion monitoring system, the response time for public opinion has been shortened from 24 hours to 4 hours, and the accuracy rate of major public opinion early warning has reached 91%, effectively improving the government's emergency response capabilities.

In the project approval scenario, large models utilize technologies such as optical character recognition (OCR), rule engines, and semantic understanding to achieve intelligent review of approval materials. After the applicant submits electronic materials, the model automatically identifies the type of materials, extracts key information, compares it with approval standards, identifies missing and incorrect items, and provides real-time feedback, supporting online correction of materials. For eligible projects, approval opinions are automatically generated and pushed to the relevant approval stages, achieving "seamless approval". After the application of the project approval large model in a provincial development zone, the review time for approval materials was shortened from an average of 5 working days to 1 working day, the approval rate increased by 15%, and the satisfaction rate of enterprises reached 96%.

3.2 Scenario Innovation for Enhancing Industrial Operation Efficiency

3.2.1 Logistics time-effect prediction and scheduling

Based on a logistics digital intelligence graph constructed using a large model, Express 100 integrates multiple sources of data such as express delivery outlets, transportation routes, weather conditions, and traffic congestion. It achieves precise logistics time prediction through Graph Neural Networks (GNN) and time series prediction models. Its core technical approach involves first converting structured data (outlet addresses, transportation distances) and unstructured data (weather texts, traffic broadcasts) into unified feature vectors using data fusion technology. Then, it utilizes the Transformer model to capture temporal dependencies and predict the dwell time at each transportation node. Finally, it combines a dynamic routing algorithm to optimize transportation routes and distribution schemes. This system enables full-chain time prediction from shipment to receipt, achieving a prediction accuracy rate of 92%. It supports hourly-level precise forecasting and has a monthly call volume exceeding 100 million times. After applying this system, a certain express delivery company reduced the delivery delay rate by 28%, decreased customer complaints by 35%, and lowered transportation costs by 12%.

3.2.2 Operation of "Thousands of Stores, Thousands of Faces" in Retail

Watsons has achieved a transformation from "standardization" to "personalization" in retail operations through its private domain operation and intelligent product selection system built on a large model. In the private domain operation segment, the model constructs multi-dimensional customer personas by analyzing customer consumption records, browsing behaviors, social interactions, and other data, enabling customer segmentation and interest prediction. Based on customer personas, personalized product recommendations, promotional activities, and other content are pushed out. Intelligent customer service responds to customer inquiries through multiple rounds of dialogues, with response time compressed from hours to seconds, resulting in a 23% increase in customer conversion rate. In the intelligent product selection segment, the model integrates LBS business district data, consumption trend data, and competitive product data, and uses association rule mining, demand forecasting, and other techniques to customize differentiated product pallets for stores in different business districts. After applying the personalized product selection scheme to core business district stores in a first-tier city, sales increased by 54% and inventory turnover rate increased by 38%. Community stores, by adapting to the consumption preferences of surrounding residents, saw a 26% increase in sales, effectively breaking the operational dilemma of the retail industry where "every store looks the same".

3.3 Innovation in Industrial and People's Livelihood Service Scenarios

3.3.1 Industrial predictive maintenance

Guangya Mingdao's industrial large model achieves predictive maintenance for industrial equipment based on the technical process of "data collection - feature extraction - model training - inference decision-making". Sensors deployed on the equipment collect operational data such as vibration, temperature, and pressure, which are preprocessed by edge computing nodes and uploaded to the cloud-based large model. The model identifies precursor features of equipment failure through time series analysis and anomaly detection techniques, and combines equipment maintenance records and expert experience to construct a fault diagnosis knowledge base, enabling fault type identification and remaining life prediction. Users can ask questions to the model through natural language interaction, such as "Possible causes of equipment vibration anomalies" and "Estimated remaining operating time", and the model quickly outputs a visual analysis report and maintenance suggestions. This system has been implemented in industries such as automobile manufacturing and machining, reducing equipment downtime due to failures by 40%, maintenance costs by 30%, and spare parts inventory costs by 25%.

3.3.2 Medical Assistance and Intelligent Customer Service

The Dr.GPT large model, which is easy and healthy to use, integrates public medical datasets (such as PubMed and MIMIC-III) with private clinical data. Through domain-adaptive pre-training and fine-tuning, it has established a medical service system encompassing intelligent consultation, chronic disease management, and claim review. In the intelligent consultation phase, the model gathers information such as patient symptoms, medical history, and medication use through multiple rounds of dialogue. It then conducts a preliminary diagnosis by combining this information with a medical knowledge graph, recommends appropriate consultation departments, provides home care suggestions for common diseases, and achieves an accuracy rate of 83% in consultation. In the chronic disease management phase, the model regularly sends medication reminders, dietary suggestions, and review notifications, monitors patients' health data in real time, provides timely alerts for abnormalities, and effectively enhances the control rate of chronic diseases.

Large model intelligent customer service has been widely applied in industries such as finance, e-commerce, and healthcare, achieving rapid response to customer needs through multimodal interaction (text, voice, image). Its core technologies include intent recognition, entity extraction, and dialogue management, which support precise answers to complex questions by constructing industry knowledge bases and dialogue template libraries. The intelligent customer service of a certain Internet healthcare platform has achieved an independent resolution rate of 85%, with only 15% of complex issues being transferred to manual processing. This has reduced the cost of manual customer service by 62% and increased customer satisfaction to 94%.

4. Industrial Empowerment Mechanism and Value of Large Model-Driven AI Implementation

4.1 Micro-Enterprise Level: Cost Reduction, Efficiency Improvement, and Decision Optimization

4.1.1 Cost structure optimization

Large models significantly reduce labor and operational costs for enterprises by replacing manual operations with automated processes. In the field of customer service, intelligent customer service can provide uninterrupted service 24/7, replacing over 60% of manual customer service tasks. After implementing intelligent customer service, an e-commerce company reduced its annual customer service costs by 58%, equivalent to saving over 20 million yuan in labor costs. In the field of data processing, large models automatically complete tasks such as data cleaning, analysis, and report generation, increasing data processing efficiency by 80%. A consulting company reduced its data analysis costs by 45%. Furthermore, large models reduce enterprise operational costs by optimizing resource allocation and reducing waste. For example, logistics companies have reduced the empty-run rate of transport vehicles by 18% through time-based prediction and intelligent scheduling; manufacturing companies have reduced production losses caused by equipment failures through predictive maintenance.

4.1.2 Improvement of decision-making efficiency

Traditional enterprise decision-making relies on manual data analysis, which poses issues such as data discontinuity and decision-making delays. The DataAgent automatic decision-making system, built with large models, integrates internal and external enterprise data to achieve full-process automation from data collection, analysis to decision execution. The system adopts a closed-loop architecture of "perception - cognition - decision-making - execution". The perception layer is responsible for multi-source data collection and integration, the cognition layer conducts data insights and trend predictions through large models, the decision-making layer generates optimal decision-making schemes based on reinforcement learning, and the execution layer interfaces with enterprise business systems through API interfaces to automatically execute decision instructions. After applying this system, a retail enterprise shortened the decision-making cycle for new product launches from 3 months to 1 month, with a 32% increase in decision accuracy; a manufacturing enterprise optimized production plans through the automatic decision-making system, reducing the order delivery cycle by 25% and increasing customer satisfaction by 20%.

4.2 Meso-Level Industry: Business Format Upgrading and Mode Reshaping

4.2.1 Industrial sector: From equipment operation and maintenance to smart factory

Large models drive the transformation of the industrial sector from traditional equipment operation and maintenance to smart factories, establishing an intelligent production system of "data - model - decision-making". In the manufacturing process, large models optimize production process parameters, enhancing product quality and production efficiency. For example, an automotive parts company optimized its injection molding process through large models, resulting in an increase in product pass rate from 92% to 98% and a 15% increase in production efficiency. In supply chain management, large models predict raw material demand and price fluctuations, optimizing procurement plans and inventory management. For instance, an electronics manufacturing company saw a 30% increase in inventory turnover rate and a 12% reduction in procurement costs. In safety management, large models identify potential safety hazards through video surveillance analysis and environmental data monitoring, leading to a 60% decrease in the incidence of safety accidents in a chemical enterprise. Platforms such as Haier Kaos and Midea Industrial Internet have developed mature smart factory solutions, driving the industrial sector towards digitalization and intelligent transformation.

4.2.2 Retail industry: from traditional offline to O+O omni-channel

Large models break down the channel barriers between online and offline retail industries, establishing an O+O (Online to Offline) omnichannel operation model of "online traffic attraction - offline experience - private domain accumulation". Online, personalized recommendations and intelligent marketing are used to enhance conversion rates. After implementing large model recommendation algorithms on an e-commerce platform, the click-through rate of products increased by 40%, and the repurchase rate increased by 28%.

Offline, intelligent shopping guides and personalized product selection improve the consumer experience. A supermarket chain analyzed consumer preferences around its stores using large models, adjusted product display and category structure, and increased store foot traffic by 18%. In private domains, community operations and precise reach enhance user stickiness. A beauty brand saw a 50% increase in average monthly consumption by private domain users. The omnichannel operation model enables retail enterprises to achieve three-level growth in "traffic - sales - loyalty", significantly enhancing the overall digitalization level of the industry.

4.2.3 Financial industry: from risk management and control to intelligent services

Large models are driving the financial industry's transformation from traditional risk management and control to intelligent services, establishing an intelligent system encompassing credit, wealth management, compliance, and other domains. In credit operations, large models assess borrower credit risk through multi-source data analysis, reducing credit approval time from days to minutes and lowering the non-performing loan ratio of a certain bank by 0.8 percentage points. In wealth management, intelligent investment advisors recommend personalized wealth management plans based on customers' risk preferences and financial status, leading to a 35% increase in the asset size of wealth management clients of a certain securities company. In compliance operations, large models automatically identify abnormal behaviors and irregularities in financial transactions, boosting the fraud claim recognition rate of a certain insurance company by 40%. The intelligent transformation of the financial industry not only enhances service efficiency and quality but also reduces operational risks, driving high-quality development in the industry.

4.3 Macro-Industry Level: Ecological Synergy and Accelerated Transformation

The large model promotes the intelligent transformation of traditional industries by building a collaborative ecosystem of "data - technology - scenario", injecting new impetus into the growth of the digital economy. Wanlian Moore's full-industry large model covers 97 major industry categories, relying on the industrial Internet platform to integrate upstream and downstream data of the industrial chain, build industry knowledge graphs and dynamic databases, and achieve continuous optimization of the model through incremental training. This model provides services such as raw material demand forecasting and supplier matching for the upstream of the industrial chain, production optimization and quality control for the midstream, and market forecasting and channel expansion for the downstream, driving an increase in the collaborative efficiency of the industrial chain by over 25%.

From an economic value perspective, large models can contribute 20% of the incremental value to the digital economy, primarily through enhanced production efficiency brought about by technological innovation and economic growth fostered by the emergence of new business models. They can also bring 80% of the incremental value to the transformation of traditional industries, mainly through efficiency improvements and cost reductions achieved after the digital transformation of these industries. According to estimates, by 2025, large models will drive China's GDP growth by over 1 trillion yuan, becoming the core engine propelling high-quality economic development. Simultaneously, large models foster cross-industry collaborative innovation. For instance, the integration of industry and finance gives rise to new supply chain finance business models, while the convergence of healthcare and technology propels the advancement of telemedicine, shaping an industrial ecosystem characterized by "cross-sector integration and collaborative win-win outcomes.".

5. Challenges and Future Outlook of Large Model-Driven AI Implementation

5.1 Existing Core Challenges

5.1.1 Technical and compliance challenges

At the technical level, large models still face issues such as hallucinations, insufficient interpretability, and high inference costs. Model hallucinations manifest as generating incorrect facts and logically contradictory content, which can lead to serious consequences in key areas such as healthcare and finance. Insufficient interpretability makes it difficult to trace the decision-making process of the model, failing to meet compliance requirements. Large model training and inference require substantial computing power support, with the training cost of a single model exceeding tens of millions of yuan, which is unaffordable for small and medium-sized enterprises. At the compliance level, risks such as training data poisoning, privacy leakage, and inconsistent technical standards are prominent. Some enterprises use unauthorized data to train models to reduce costs, posing intellectual property infringement risks. When large models handle sensitive data, privacy leakage incidents are prone to occur. The industry lacks a unified technical standard and evaluation system, making it difficult to compare the performance of different models and affecting large-scale applications.

5.1.2 Challenges in scenario and industry adaptation

In terms of scenario adaptation, large model applications face issues such as severe homogenization and a disconnect between technology and business. Most enterprises blindly follow the trend in deploying general-purpose large model applications, lacking in-depth exploration of vertical industry needs, resulting in single application scenarios and limited empowerment effects. Some applications only remain superficial, failing to deeply integrate with enterprise business processes, making it difficult to achieve value transformation. At the industrial level, issues such as insufficient industrial chain collaboration and high data barriers hinder the implementation of large models. The data standards of upstream and downstream enterprises in the industrial chain are not unified, making it difficult to achieve data sharing. Some enterprises, out of data security concerns, are reluctant to open up core data, leading to low-quality training data for large models and affecting model performance. The shortage of AI composite talents is prominent. According to data from a recruitment platform, the supply-demand ratio of AI talents stands at 1:10, and the shortage of composite talents who understand both technology and the industry is even greater, constraining the ability of enterprises to implement large models.

5.2 Future Development Trends

5.2.1 Technological development trends

Large model technology will evolve deeply towards lightweight, vertical, and trusted directions. In terms of lightweight, model compression technology will continue to be optimized, and breakthroughs in hardware technologies such as quantum computing and new chips will further reduce inference costs, making large models on the edge an important development direction. In terms of verticalization, industry-specific large models will focus on the needs of specific domains, enhancing adaptability through domain-adaptive training. For example, medical large models and industrial large models will become research and development priorities. In terms of trustworthiness, technologies such as explainable AI, privacy computation, and security detection will continue to be improved, forming a standardized trusted technology system to provide security guarantees for the implementation of large models. Furthermore, the integration of large models with technologies such as robotics, the Internet of Things, and blockchain will become more profound, building a full-chain intelligent system of "perception - cognition - decision-making - execution".

5.2.2 Scenario and industry trends

Scenario applications will deepen from general to specialized, with more high-value vertical scenarios being explored. In the industrial field, large models in material science will drive the research and development of new materials, while large models in spatial intelligence will optimize factory layouts and logistics paths. In the medical field, large models in genetics will facilitate precision medicine, and large models in drug research and development will shorten the cycle of new drug development. In the educational field, large models in personalized tutoring will enable teaching according to students' aptitudes, and large models in vocational education will enhance the effectiveness of skill training. At the industrial level, a technology standard system guided by policies will be gradually established, with industry associations and leading enterprises leading the development of technical specifications, safety standards, and evaluation methods for large models, promoting orderly development of the industry. Industrial chain collaboration will be further strengthened, forming a collaborative ecosystem of "data providers - technology service providers - scenario application providers", with mechanisms for data sharing and computing power sharing continuously improved. Large models will upgrade from single-point applications to ecological collaboration, driving traditional industries to achieve comprehensive and deep digital transformation, ultimately moving towards a new stage of "human-machine symbiosis".

References

[1] Liu Shiqi, Liu Zhi, Duan Huimin, et al. --Technical Architecture, Development Status, Practical Paths, and Future Outlook [J]. Journal of Distance Education, 2025, 43(1):33-45.

[2] Yin Ximing, Wu Peiqi, Qian Yating, et al. Scene-driven digital and intelligent technology innovation empowers new productive forces: theoretical logic and practical approach [J]. China Soft Science, 2024(10):18-31.

[3] Fan Yuan. Research on the Value Realization of 5G Technology for Industrial Internet [D]. Beijing University of Posts and Telecommunications, 2022.

[4] Feng Xia, Mao Lingfeng, Xu Tingting, et al. Overview of Large Language Model Security Research for Vehicle-Road-Cloud Cooperative Autonomous Driving [J]. Journal of Communications, 2025, 46(11).

