Operator Best Practices

AI Large Model Empowering Verticals Use Cases

Industrial Upgrading & Intelligent Manufacturing | Public Services & Social Governance | Healthcare & Smart Education | Innovative Customer Services & Operations

2025

GSMA™

Foreword

Sihan Bo CHEN
Head of Greater China
GSMA

In 2024, GSMA Greater China released the "Operator Practice: 2024 Benchmark Cases of AI Large Model-Empowered Verticals," presenting typical cases in fields such as government services, customer services, industrial applications, healthcare, education, cultural tourism, and urban governance from operators' perspectives. Focusing on the innovative applications and exploratory practices of large AI models across multiple industries, this report demonstrates how technology is breaking through industrial boundaries, reveals the challenges behind success and the value of innovation, and offers profound insights and Chinese expertise to global operators and vertical industries.

Entering 2025, the industry landscape is evolving at an accelerated pace. The deep integration of large AI models with 5G-Advanced, cloud computing, and computing power networks has been a key driving force for digital economic development. This year, the practices of Chinese operators have presented marked progress and transformation compared to last year: the case studies no longer merely emphasize the breadth of applications and exploratory pathways, but instead focus more on replicable, measurable, and iterative industrial implementation. The convergence of cloud-network-intelligent computing, industry-specific large models, and agent platforms has become mainstream. Model capabilities have moved beyond single-point tools, and a value-closed loop integrating training and inference has been formed through engineering methods and iterative data feedback. Operators and ecosystem partners are co-innovating to advance AI capabilities from pilot projects to systematic deployment at scale, truly translating technological bonuses into long-term value for society and industry.

This year's case studies collection concentrates on four key areas: innovative customer services and operations, healthcare and smart education, industrial upgrading and intelligent manufacturing, and public services and social governance. They systematically outline the pathways and outcomes of large AI models, empowering digital transformation across sectors. These cases not only cover a wider range of industries but also emphasize quantifiable results and clearer business models. For instance, China Telecom's education platform serves over 2,000 schools in 31 provinces, increasing grading efficiency by 70% with over 95% consistency; China Unicom's healthcare quality control solution has covered petabyte-level data volumes of 230 institutions. Millions of daily quality checks are conducted, shortening the quality control cycle from "years" to "months," and saving nearly RMB10 million in medical insurance costs; China Mobile's civil aviation customer service solution has reduced order processing time by 40%, increased recognition rates for high-frequency business scenarios by 95%, and cut agent interaction time by over 50%. Operator-specific capabilities like Open Gateway and cloud-network integration have also been incorporated into solution architectures, promoting ecosystem collaboration with governments, hospitals, enterprises, and AI partners, and moving from pilot phases to routine operations.

These practices suggest that Chinese operators shifted their focus from "application breadth" to "implementation depth and industrial value reconstruction." They are not only driving intelligent industrial upgrading and unleashing the value of data factors but also playing an indispensable role in promoting high-quality social development. Through engineering methods and deep integration with business scenarios, large model technology forms a sustainable innovation closed-loop, bringing replicable, measurable, and iterative implementation models to industries, and bringing the development of Digital China to new heights.

At the same time, responsibility and security are the foundation of technological advancement. As large AI models penetrate deeper into various sectors, challenges such as data security, privacy protection, model generalization, and integration of domain knowledge persist. Achieving deep synergy between technology and business models, and building open, win-win ecosystems remain areas requiring continuous effort. Besides long-term investment and strategic resolve, operators and ecosystem partners need to jointly advance AI, within a framework of open cooperation and responsible innovation, to support sustainable economic and social development.

Looking ahead, large AI models will continue to evolve, and application scenarios will expand. Operators will unleash innovative vitality across more domains, joining hands with ecosystem partners to co-create a new intelligent era. We believe that large AI models will play significant roles in more industries, driving digital transformation and upgrading, fostering new applications and business models, and creating greater value for industries and society at large.

This case collection is not only a systematic summary of Chinese operators' innovative practices with large AI models but also a sharing of experiences with global industry peers. By sharing these benchmark cases, we want to inspire more innovative practices that can contribute to digital transformation and high-quality development in China and beyond.

Finally, we extend our sincere gratitude to all the teams and experts involved in the compilation and publication of this case collection, as well as to the China Artificial Intelligence Industry Development Alliance (AIIA) for their close collaboration. Let us join hands and stride forward together to embrace a future where large AI models empower industries and society!

Contents

Industrial Upgrading & Intelligent Manufacturing

Kunlun Large Model Empowers Energy and Chemical Industry Transformation 03

Joining Forces to Build an AI + Agriculture Large Model Project 07

Unicom Ship Design Copilot — Large Model-Driven Intelligent Transformation of Ship Design 11

Public Services & Social Governance

AI-Powered Environmental Protection: Practice of the Xingchen Eco-Environment Governance Large Language Model 17

TeleNavi Xingchen Wendao Large Language Model Supports Intelligent Mobility for Seniors 21

Large Model-Empowered Human Resources and Social Security Inquiry 27

Telecom LLM-enabled Digital Intelligence Products 32

Healthcare & Smart Education

Large Model Project for Medical Imaging Quality Control 36

Building Smart Hospitals Based on Medical Large Models 40

Innovative Practice of Large-Scale Personalized Education Utilizing Education-Specific Large Models 44

Innovative Customer Services & Operations

Civil Aviation Customer Service Large Model and Intelligent Applications 49

Harnessing AI to Reshape Enterprise Training Efficiency 53

"Smart Clothing": AI-Driven Digital Relics and Immersive Cultural 56

Kunlun Large Model Empowers Energy and Chemical Industry Transformation

> **The deep integration of artificial intelligence with the energy sector has become a pivotal force in driving industrial transformation, resource reallocation, and competitive advantage restructuring. As central enterprises, China Mobile and CNPC (China National Petroleum Corporation) are leading the charge to enhance R&D, production, operations, and service management in the energy-chemical sector through AI. This initiative synergizes "Traditional Industries + AI" with "AI + Emerging Industries" for collaborative growth.**
>
> **Yang Wenjun**
> **Director of CNPC Digital Intelligence Research Institute**

SOLUTION PARTNERS

The wave of artificial intelligence is sweeping in; technology is iterating rapidly and has become the focal point of great-power technological strategic competition. General Secretary Xi Jinping has emphasized that, faced with the new situation of the rapid evolution of a new generation of artificial intelligence technologies, we must fully leverage the advantages of a new type of whole-of-nation system, insist on self-reliance and self-improvement, emphasize application orientation, and promote the healthy and orderly development of China's artificial intelligence toward being beneficial, safe, and fair. The 20th National Congress of the Communist Party proposed promoting cluster development and integration of strategic emerging industries, and building a number of new growth engines such as next-generation information technology and artificial intelligence.

Since the high-quality advancement of the Kunlun large model began in 2024, its construction has adhered to winning by quality, accelerated the pace to seize opportunities, and deeply empowered business areas across the energy and chemical industries. Relying on deep technical accumulation in artificial intelligence and rich experience in informationization project implementation, China Mobile built the "China Mobile Energy & Chemical Industry Large Model," providing full-business, full-process support to the Kunlun large model construction. It has worked with CNPC to explore paths and methods for innovative applications of artificial intelligence in the petroleum and energy sectors, fully supported the construction of "Digital & Intelligent CNPC," and advanced AI-enabled new industrialization of the national energy and chemical industries. Within one year of project construction, China Mobile has built a three-tier computing network of cloud-edge-device, developed an 80-billion-parameter multimodal industry large model and tackled forward-looking technologies such as multimodal capability fusion architecture, pioneered a unified AI middle platform for the energy and chemical industry, deployed 12 high-value scenarios, and empowered core businesses including exploration & development, refining & production, and design services. This comprehensively assists CNPC in improving quality and efficiency in production, safety, management, and service across its fields.

CHALLENGES

Insufficient computing resources, unable to meet model training and deep application demands

When building an industry large model, computing power determines model training speed and efficiency, and affects the performance of the industry large model the group builds, the efficiency of experimentation and exploration, and system real-time performance and responsiveness. At present, most enterprises are clearly short of computing resources for intelligent computing and cannot meet AI construction needs; high-quality computing support is urgently required.

Fragmented data resources, difficult to effectively support digital transformation

Currently, most enterprises' corpora are dispersed and isolated. The majority of unstructured data is scattered across the group's various companies and on individual computers. A coordinated corpus collection system needs to be established, and continuous collection work should be carried out during project implementation.

Small models have weak generalization capabilities, making them difficult to replicate and scale

The energy and chemical sector has a long industrial chain and complex businesses. Existing small models and applications built on small models often offer single functions and weak generalization, making it difficult to meet complex business needs across different scenarios. It is necessary to build a unified industry large model and combine large and small models to balance specialization and replicability.

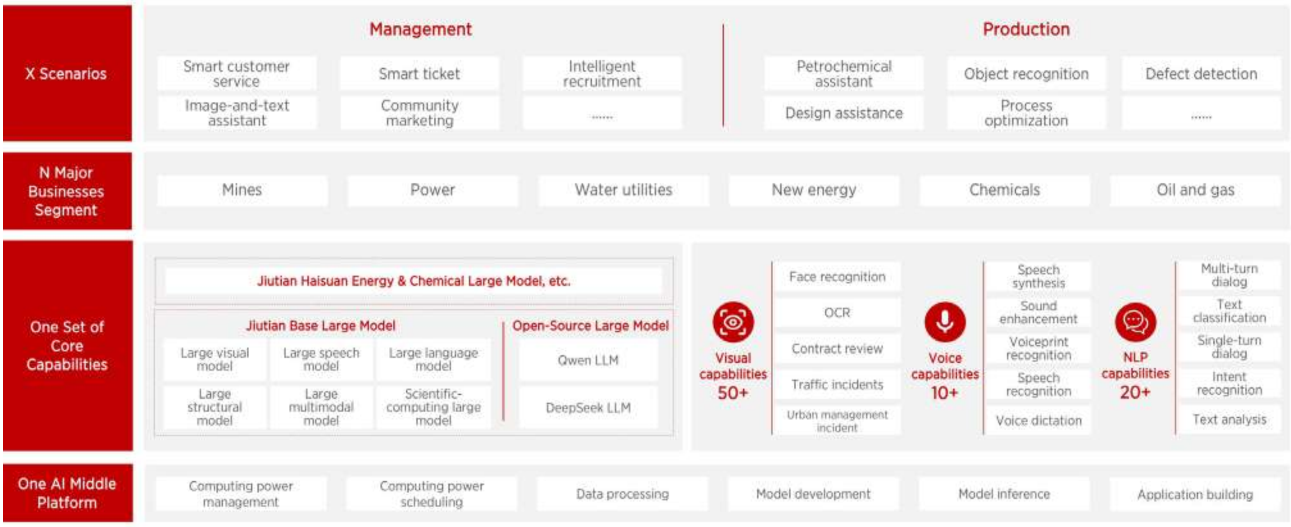
Applications are not systematized; there is a lack of unified top-level planning and standards

AI technology is evolving rapidly. Current AI+ application scenarios are mostly developed as localized, point solutions, lacking group-level unified planning and relevant standards. A business blueprint and systematic advancement approach have yet to be formed, making it hard to generate effective synergy and hampering enterprise digital transformation.

SOLUTION AND VALUE

01 Overall Structure/Solution

China Mobile and CNPC have built an AI construction system for the full energy-and-chemicals business chain under the "1+1+N+X" model. This includes a unified AI middle platform that covers compute resource governance, data processing, model development, and application construction; a coordinated core capability comprised of one set of industry "large model + small models"; services for N major energy business domains; and X categories of typical scenarios defined across management and production domains to effectively advance intelligence in the energy and chemical industry.



X Scenarios

Management: Smart customer service; Smart ticket; Intelligent recruitment; Image-and-text assistant; Community marketing; ......

Production: Petrochemical assistant; Object recognition; Defect detection; Design assistance; Process optimization; ......

N Major Businesses Segment

Mines; Power; Water utilities; New energy; Chemicals; Oil and gas

One Set of Core Capabilities

Jiutian Haisuan Energy & Chemical Large Model, etc.

Jiutian Base Large Model: Large visual model; Large speech model; Large language model; Large structural model; Large multimodal model; Scientific-computing large model

Open-Source Large Model: Qwen LLM; DeepSeek LLM

Visual capabilities 50+: Face recognition; OCR; Contract review; Traffic incidents; Urban management incident

Voice capabilities 10+: Speech synthesis; Sound enhancement; Voiceprint recognition; Speech recognition; Voice dictation

NLP capabilities 20+: Multi-turn dialog; Text classification; Single-turn dialog; Intent recognition; Text analysis

One AI Middle Platform

Computing power management; Computing power scheduling; Data processing; Model development; Model inference; Application building

Figure 1: China Mobile's energy and chemical industry construction system

02 Application Scenarios

Sales and customer service for oil & gas products

In oil-and-gas product sales, there are large numbers of both business and individual customers; common problems include poor convenience and slow response, and customer-service system capabilities and staff skill levels vary widely. For the 956100 customer service system, in multi-turn conversations with complex expressions and ambiguous semantics, the large model precisely recognizes customer intent and uses more natural expressions to provide customers with smooth, professional, and human-like conversational experiences. In question-answer assistance scenarios, conversation summarization and knowledge recommendation are implemented, improving service efficiency by 40%. In ticket filling, the large model automatically summarizes conversation content and helps agents create tickets with one click. Additionally, the large model can generalize customer-service knowledge at scale to improve operational efficiency.

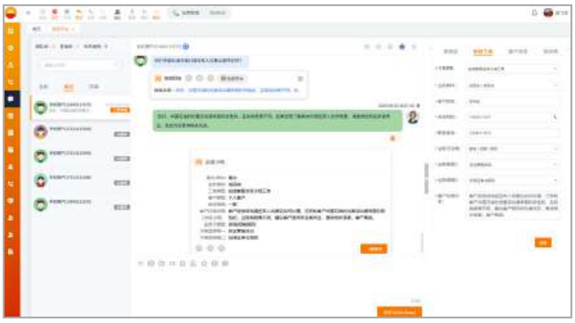
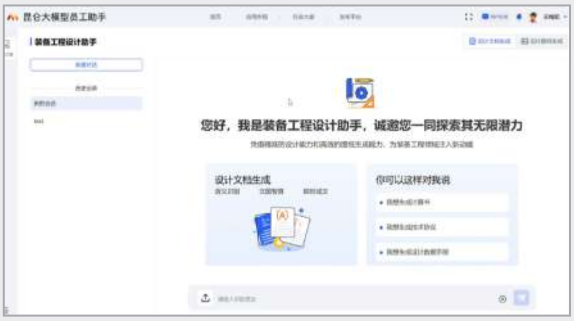


Figure 2: Intelligent customer service interface



您好，我是装备工程设计助手，诚邀您一同探索其无限潜力

Figure 3: Equipment engineering design assistant interface

Engineering design

Based on China Mobile's engineering-design large model core capabilities, a joint equipment engineering design assistant was built to generate hydraulic-control design documents. Compilation efficiency improved from days to minutes, supporting more efficient production. Innovative capabilities to generate equipment manufacturing drawings enable end-to-end one-click output of CAD drawings for top-drive electrical control rooms, promoting intelligent design.

Production

On-site CDU (atmospheric and vacuum distillation unit) process data, images, and other information are processed by the multimodal large model to provide intelligent operation assistance, operation and training support, and instant answers to questions about process technology, unit operation, accident plans, and start/stop procedures. Model fidelity accuracy exceeds 95.5%; using a structured-prediction large model during unit operation, future trends of process parameters are forecast. Through anomaly detection and root-cause localization, the system alerts operators to possible operational fluctuations and process errors, ensuring safe and stable CDU operation and reducing energy consumption.

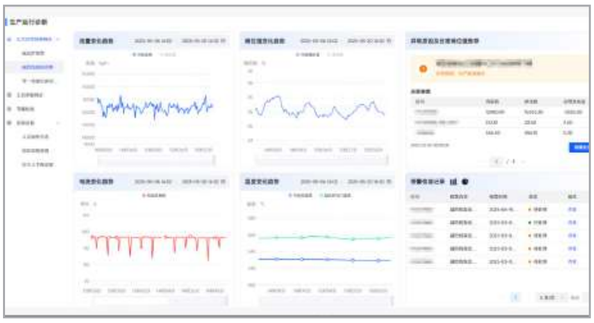


Figure 4: Frontline liquid-level fluctuation early-warning interface

昆仑大模型员工助手
社群营销

Figure 5: Community marketing interface

Community Marketing

Build a "Marketing Consultant Assistant" robot based on WeChat Work, leveraging AI capabilities such as user inquiry intent recognition, key parameter extraction, NL2SQL conversion, and natural language generation from large language models. This robot provides end-to-end support for marketing campaigns, encompassing diverse functions like real-time campaign consultation responses, precise campaign information dissemination, personalized marketing copy generation, and emotional good morning greetings. With the goal of "From morning warmth to viral copywriting—AI marketing strategist is always online," it offers new productivity enhancements for WeChat Work community operations, empowering over 20,000 gas stations and 100,000+ community operations nationwide.

03 Key Innovations

Multimodal data fusion

China Mobile and CNPC's Kunlun large model innovatively builds a multimodal data fusion system suitable for the complex environment of the energy and chemical industry. The system efficiently integrates geological images, equipment operational audio, production-process video, and textual process descriptions and equipment parameters.

Breakthroughs in cross-modal understanding and generation

Kunlun large model has made key advances in cross-modal understanding and generation. First, amazing cross-modal understanding. Given, for example, an image of a faulty refining device, the model not only recognizes the fault type from the image but also retrieves and analyzes textual knowledge to understand the fault's impact across the refining process and potential chain reactions. Second, excellent generation. From a textual description of a new oil-and-gas field development plan, the model can generate corresponding multimodal outputs such as visualized geological models and development planning flowcharts. This cross-modal generation capability greatly improves communication and decision efficiency, enabling personnel from different professional backgrounds to better understand complex energy projects and facilitating cross-department collaboration.

Multimodal scene-adaptive optimization

To address the industry's diverse business scenarios, Kunlun implements multimodal scene-adaptive optimization to achieve maximal performance across different contexts, significantly improving multimodal techniques' practicality and reliability in real energy production and operations.

04 Business Models

Kunlun large model positions itself as "full-chain empowerment for the energy industry" and establishes a three-dimensional commercial model covering standard products, R&D, implementation management, and operation & maintenance.

Standard products	R&D
Offer standardized models, middle platform, and compute products/services that cover common energy scenarios and support base-level intelligent applications, charged according to standard pricing	Focus on customized scenarios and provide full lifecycle services from design and development to model training and fine-tuning, charged by workload to address industry-specific intelligent needs
Implementation management	**Operation & maintenance**
Provide overall project integration and management services (progress control, stage reviews, deliverable acceptance) to ensure project landing quality, charged by workload	Provide basic and value-added O&M services to keep systems stable, offer operational services, and continuously update knowledge, data, and next-generation technology to meet long-term intelligent operational needs, charged by workload

05 Core Values

Using the "China Mobile Energy & Chemical Industry Large Model" as the technical foundation, the "1+1+N+X" model realizes comprehensive empowerment of CNPC's businesses and provides a benchmark case for central SOEs' digital transformation and AI+ energy and chemical industry adoption. The Jiutian AI middle platform hosts over 80 model training and inference tasks, enables unified management of models from multiple vendors, elastic scheduling of compute resources, and end-to-end dataset management. It lays a solid foundation for CNPC to aggregate innovation factors, foster diverse intelligent applications, and build a flourishing AI technology ecosystem. The energy large model powers high-value energy-and-chemical-specific applications such as intelligent CDU operation and diagnosis, object recognition, community marketing, and recruitment assistants. These applications have driven changes in production operations, marketing, and corporate management, producing tangible business transformation results and helping to establish an initial ecosystem for AI-driven innovation.

REFLECTION

Key Insights

China Mobile's energy & chemical industry large model has efficiently supported the construction of the Kunlun large model, driving continuous progress and observable results in AI-enabled industrial transformation. It has provided a reproducible and successful reference for the deep integration of artificial intelligence with traditional industries.

Future Plans

Moving forward, China Mobile will continue to leverage its strengths in technological innovation, explore paths and methods for innovative AI applications in the energy and chemical sectors, fully support the construction of a "Digital & Intelligent CNPC," and further advance AI-enabled new industrialization of the national petroleum industry; build an autonomous and controllable AI technology high ground. Focusing on the national "AI+" strategy and energy security needs, establish a dual-engine approach of "jointly tackling core technologies + co-building a domestic ecosystem," tackle the fundamental technologies of energy-industry large models, and construct a full-stack, independently controlled core capability base; deepen thoroughly empowering AI application scenarios and concentrate on CNPC's core businesses — oil & gas and new energy, refining & new materials, engineering services, and capital & finance — and use AI to restructure value across the entire industry chain, achieving full-link, full-scenario integration from "geological exploration and engineering design to production & operation, supply-chain management, and market sales"; build an open and shared AI industry ecosystem. With the aim of "empowering the industry and serving the whole," co-develop industry standards and open platforms to advance the energy-AI ecosystem from isolated "single-point breakthroughs" to a "systematic and flourishing" ecosystem.

Joining Forces to Build an AI + Agriculture Large Model Project

> Ensuring food security is one of the core tasks of the national strategy. At present, China's grain production and storage still face prominent problems such as strong pressure to increase yield per unit area, low levels of informatization, and weak risk early warning capabilities. There is an urgent need to drive agriculture toward intelligent and precise transformation through technological innovation. China Mobile, together with leading enterprises, has built the agricultural large model. It can provide industry-wide intelligent Q&A services for agricultural operations and also focus on specific agricultural scenarios by integrating smart agricultural facilities and professional agronomic solutions to deliver precise production and management guidance.

Chen Zhiyong
"Ten-Hundred-Thousand" Agricultural Large-Model Expert
China Mobile (Chengdu) Industry Research Institute

SOLUTION PARTNERS

Building on its core capabilities in cloud networking and innovative AI large model algorithms, China Mobile has forged a strategic partnership with two leading industry giants—Sinochem Group and the State Grain Reserve Corporation—to develop agricultural industry large models tailored to two specific scenarios: crop cultivation and grain storage. For the crop cultivation scenario, the model incorporates functions such as agronomic Q&A, pest and disease identification, and water and fertilizer recommendations. By integrating advanced technologies like knowledge graph retrieval-augmented generation and locality-specific prior knowledge parsing, it delivers more precise and professional planting guidance. The outcomes of this scenario have been applied to Sinochem's iMAP product. Meanwhile, the grain storage scenario focuses on meeting demands such as quantifying grain storage processes, monitoring, early warning, and prevention of harmful organisms, as well as enhancing the skills of personnel. It provides granary managers with intelligent operational guidance and decision support, with the results integrated into the Jiyuan Grain Storage Large Model.

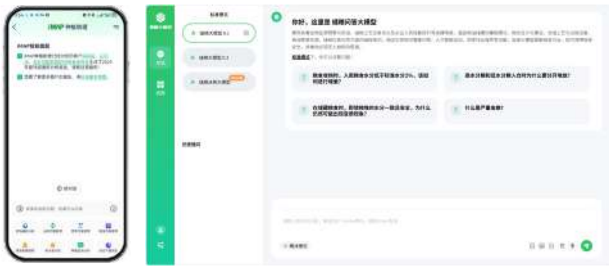


Figure 1: agricultural large model applications: cultivation (left) and grain storage (right)

CHALLENGES

Traditional agricultural production and management mainly face the following challenges:

Decision making relies on experience

Planting based on "conventional lore" easily misjudges phenological stages, reduces resilience to natural disasters, and increases the risk of yield loss.

Delayed pest and disease identification

Manual inspections have high misidentification rates; missing the optimal control window can cause large yield reductions.

Inefficient resource use

Flood irrigation and indiscriminate fertilizer application waste water and lower fertilizer use efficiency.

Labor shortfall

Rural labor is aging and hollowing out; smallholders operate with low efficiency and high labor costs.

Shortage of skilled personnel

Granary professional teams are aging—veteran workers have rich experience but are slow to adopt new technologies; new hires are quicker with tools but lack practical experience. Addressing the talent gap in grain storage is an urgent industry task.

Difficulty in monitoring storage pests

Stored grain is constantly threatened by pests. Traditional monitoring methods cannot achieve full-area, real-time, precise surveillance and early warning, making timely responses to pest risks difficult and causing severe post-harvest losses.

SOLUTION AND VALUE

01 Overall Structure/Solution

China Mobile, together with Sinochem Group and the State Grain Reserve Corporation, has accumulated and consolidated large-scale, high-quality multimodal industry data covering both cultivation and grain storage management. The dataset includes agricultural practice records, professional agronomic explanations, and practical prescriptions, supporting the training of the agricultural industry large model. Advanced AI algorithms and big-data processing techniques are used to integrate, analyze, and model these agricultural knowledge resources, enabling the large model to handle complex agricultural tasks with high accuracy and reliability. Built on an advanced general large model foundation and combined with agricultural specialty small models, the system delivers practical functions across on-farm production and post-harvest stages, providing strong technical support and decision basis for agricultural producers and promoting digitization and intelligent development of the agriculture sector.

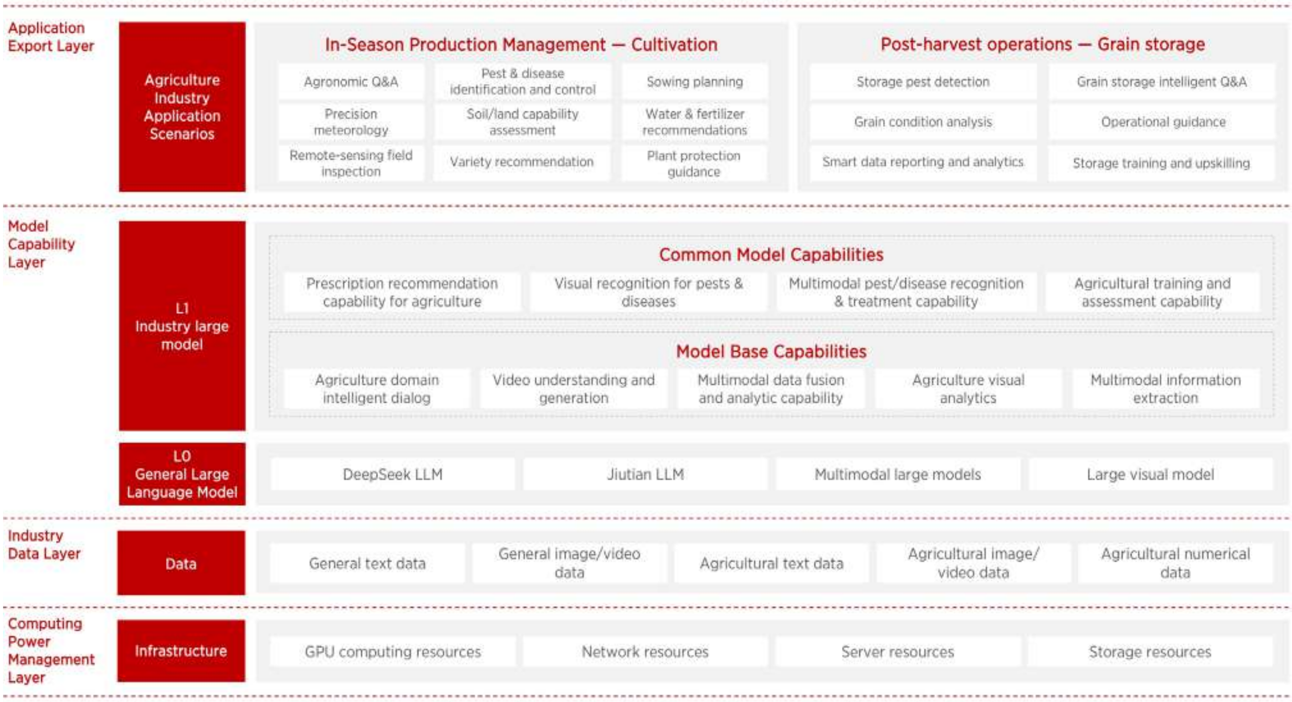

Application Export Layer
Agriculture Industry Application Scenarios
In-Season Production Management — Cultivation
Agronomic Q&A
Pest & disease identification and control
Sowing planning
Precision meteorology
Soil/land capability assessment
Water & fertilizer recommendations
Remote-sensing field inspection
Variety recommendation
Plant protection guidance
Post-harvest operations — Grain storage
Storage pest detection
Grain storage intelligent Q&A
Grain condition analysis
Operational guidance
Smart data reporting and analytics
Storage training and upskilling
Model Capability Layer
L1 Industry large model
Common Model Capabilities
Prescription recommendation capability for agriculture
Visual recognition for pests & diseases
Multimodal pest/disease recognition & treatment capability
Agricultural training and assessment capability
Model Base Capabilities
Agriculture domain intelligent dialog
Video understanding and generation
Multimodal data fusion and analytic capability
Agriculture visual analytics
Multimodal information extraction
L0 General Large Language Model
DeepSeek LLM
Jiutian LLM
Multimodal large models
Large visual model
Industry Data Layer
Data
General text data
General image/video data
Agricultural text data
Agricultural image/ video data
Agricultural numerical data
Computing Power Management Layer
Infrastructure
GPU computing resources
Network resources
Server resources
Storage resources

Figure 2: Structure of agricultural large model

02 Application Scenarios

Cultivation guidance

Leveraging a rich professional agronomy knowledge base and using plot location, remote sensing, IoT devices, and meteorological data, the agricultural large model provides crop stage-specific cultivation plans covering five major aspects: soil/land capability assessment, variety recommendation, sowing planning, water and fertilizer suggestions, and plant protection schemes.

Pest and disease identification and control

Capable of recognizing 24 crop types and 55 crop pests/diseases, the model recommends effective, low-toxicity control agents and scientifically reasonable timing and application methods to curb pest and disease spread.

Remote sensing field inspection

Using remote sensing image recognition and large model decision analysis, abnormalities in fields can be pinpointed within 5 seconds, with production suggestions and mobile navigation to the site, greatly reducing manual inspection costs.

Grain storage Q&A

Integrating policies and regulations, authoritative industry standards, and operational experience into a vast data system, the model provides fast, professional, and accurate Q&A services for grain storage practitioners.

Remote operational guidance

By integrating AR equipment for augmented-reality capabilities, the system acquires real-time data, which the large model analyzes to assist workers with granary inspections, pest control, and safety risk early warning, providing step-by-step operational guidance.

Intelligent data-driven reporting

Aggregating multi-source storage data, extracting business data and generating analytical reports, the system combines visualization and structured text to automatically produce reports. It intelligently summarizes and forecasts key indicators such as temperature, humidity, pest situation, and ventilation trends to support managers in optimizing decisions.

03 Key Innovations

Achieving precise multimodal agricultural data coordination through space-time alignment and inference using integrated satellite-ground techniques

Ground monitoring equipment offers high precision and frequency but limited coverage, while satellite remote sensing provides wide coverage with lower temporal frequency and lower resolution. Existing satellite-ground integration solutions suffer from "temporal gaps" and "spatial gaps," preventing all-around, high-precision, high-frequency monitoring. To address this, the solution proposes an integrated approach combining satellite-ground spatio-temporal association, spatio-temporal inference, and information interpretation. The system systematically builds capabilities for fusing heterogeneous satellite-ground data, joint spatio-temporal inference, and regulatory/operational analysis.

Unsupervised pest recognition based on feature-correlation selection to improve pest detection efficiency and accuracy

Pest recognition is a low-incidence, hard-to-collect scenario, so large labeled datasets are typically unavailable to support conventional supervised algorithms. The proposed solution introduces an unsupervised recognition technique driven by feature correlation. By self-learning from limited images and combining tensor-fusion methods, it delivers low-cost, high-accuracy pest identification for such special scenarios.

Knowledge-graph-driven controllable content generation for large models to enhance scientific rigor and explainability of agricultural intelligence decisions

Large model hallucinations substantially undermine model reliability and answer correctness. To mitigate this, the solution adaptively constructs a knowledge graph and cascades the knowledge graph structure into the model's information pipeline, constraining outputs to improve the trustworthiness and quality of generated content.

04 Business Models

User type	Service	Sales model
Agricultural parks and production entities	Agronomic service large model: intelligent Q&A, alert push & recommendations, agricultural technology training, guided functions, etc.	Cloud-hosted large model application services
Agricultural bureaus	Office/planning large model: intelligent Q&A, data retrieval & integration, document preparation, policy interpretation & queries, etc.	Localized large model deployment + application software
Grain storage enterprises	Grain storage large model: intelligent Q&A, data retrieval & analysis, operational guidance, grain condition forecasting, etc.	Localized large model deployment + application software

Two modes according to project needs:

- Cloud large model application services — for users with more common needs. Standard modules are charged per service module, with support for customized development and local deployment.
- Local deployment of the large model and application software — for users with stronger data privacy concerns. One-time charges apply for localized deployment, compute resources, and application software.

05 Core Values (including quantitative results and qualitative impact)

Help reduce costs and increase efficiency in agricultural production

Precision weather provides field-level forecasts, reducing losses from extreme weather; rapid pest/disease diagnosis lowers control costs and misdiagnosis risk; remote-sensing field inspection combined with intelligent decisioning optimizes fertilization and irrigation usage. By integrating agricultural resources and technologies, management becomes more precise and efficient, cutting unnecessary inputs and costs.

Demonstrated outcomes	labor costs ▼ about **50%**	agronomic decision time ▼ **75%**	per-mu water & fertilizer costs ▼ about **15%**	staple-crop per-mu yield ▲ over **10%**	**substantial productivity gains**

Support food security assurance

Consolidating authoritative grain storage data and over 1,000 industry books and standards provides practitioners with fast, professional knowledge support to improve efficiency and standardization; an open-set detection visual model for pest detection achieves accuracy above 90%, significantly improving timeliness and accuracy of storage pest monitoring; combining AR and AI enables real-time assisted operational guidance; extracting business data and generating visual analysis produces intelligent summaries and trend forecasts for key grain indicators, strengthening the scientific basis and efficiency of management decisions. These technological and intelligent applications in storage increase warehousing efficiency, reduce operating costs, facilitate knowledge transfer and talent training, accelerate industry digital transformation, and provide a robust contribution to national food security.

REFLECTION

Key Insights

Unblocked the upstream-downstream agricultural value chain and deeply integrated industry resources to promote efficient agricultural development.

Upstream cooperation	Partnering with upstream leaders (e.g., Sinochem/Syngenta Group) to share resources and complementary strengths, improving the agricultural large model's accuracy and professionalism for cultivation guidance.
Midstream application	Serving planting and production processes by fusing multisource agricultural data to deliver precise guidance and advance agricultural intelligence upgrades.
Downstream enablement	Working with the State Grain Reserve Corporation to apply the large model for intelligent warehousing management, achieving smarter, more efficient grain storage.

Future Plans

Promote scaled deployment	Consolidate lessons from pilot projects to form reproducible, scalable extension models.
Improve decision-support capability	Establish routine data update mechanisms and continually expand and refresh the model knowledge base. Upgrade the decision engine: Apply more advanced algorithms and machine learning techniques to optimize and evolve the model's decision engine.
Empower more agricultural scenarios	Expand into additional application areas such as variety breeding support, agricultural input recommendations, and price forecasting.

Unicom Ship Design Copilot — Large Model-Driven Intelligent Transformation of Ship Design

> The "Lianzhou SmartShip" project is a joint initiative by China Unicom Data Intelligence, Shanghai Unicom, and CSSC Seagosoft, focusing on three major pain points in ship design: complex regulations, inefficient modeling, and slow customization response. Built on the self-developed UniAI multimodal large model, the project offers three intelligent agent products — Regulation Q&A, Modeling Design, and Software Design — to reconstruct the entire ship design workflow using AI. Application results are significant: regulation query efficiency improved by 30%, 3D modeling labor reduced by 40%, and software design development cycles shortened by 50%. The project strongly promotes the intelligent transformation of domestically produced shipbuilding, supporting the national "dual-carbon" strategy and the development of a strong maritime nation.
>
> Ding Ding
> China Unicom Data Intelligence Co., Ltd.
> Deputy General Manager

SOLUTION PARTNERS

The ship design industry is currently at a critical stage of digital and intelligent transformation. Rising requirements in maritime transport for energy savings and emissions reduction, operational efficiency, and safety are driving ship design beyond traditional structural and economic considerations toward deep integration of environmental protection, automation, and intelligence. Technologies such as artificial intelligence, big data, and cloud computing are driving upgrades in the design process. However, the industry still faces challenges in balancing design innovation with cost control, complying with complex technical regulations, and achieving standardization.

Against this backdrop, the "Lianzhou SmartShip: AI Hub for the Shipbuilding Industry" project was launched. The project aims to integrate advanced algorithms with ship design scenario data and leverage core AI technology — the UniAI multimodal large model — to precisely address core pain points such as complex regulatory requirements, long modeling and design cycles, and slow customization response. The project has built an end-to-end solution covering regulation Q&A, modeling design, and software system design assistance. Relying on agent collaboration, toolchain support, and an algorithm engine, it reshapes the data-driven productivity of ship design.

The project seeks to form standardized products for scalable promotion, closing the "pain point to solution" loop to significantly improve design efficiency and unlock data value. For example, the "Modeling Design Agent" can efficiently identify parameters and elements from 2D drawings and automatically output 3D modeling data, greatly improving efficiency and reducing overall project cost.

The project carries multiple strategic significances

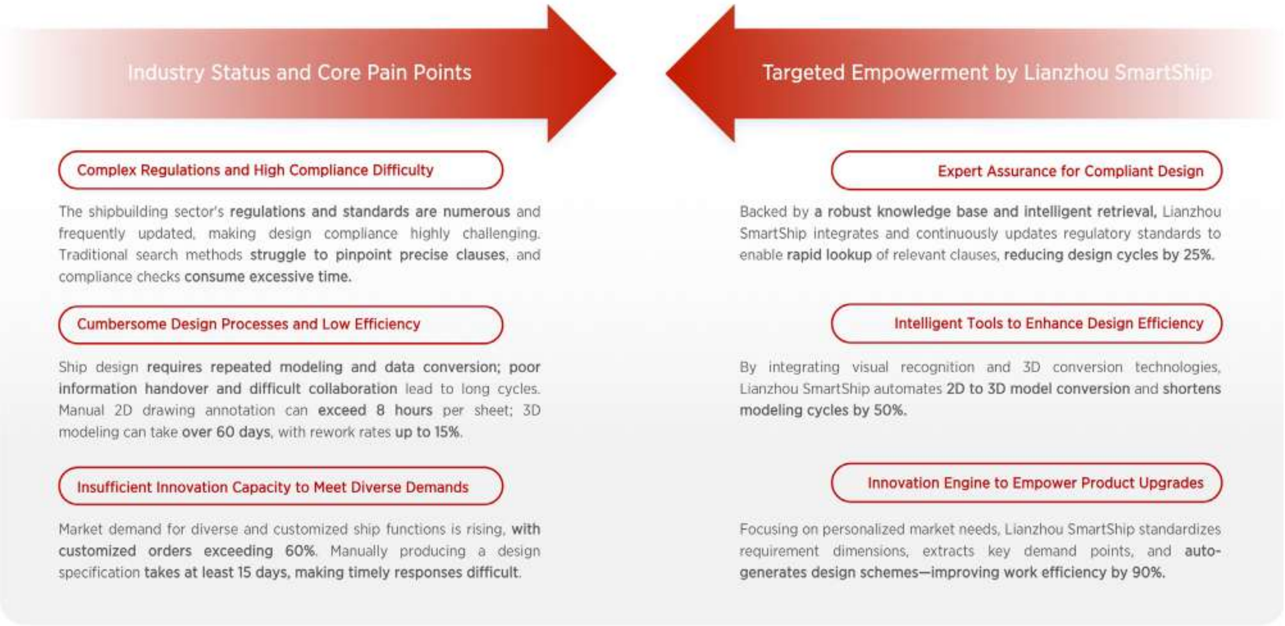
01 strengthening the global competitiveness of China's shipbuilding industry

02 providing a benchmark for industry digital transformation and leading improvements in intelligent design capabilities

03 optimizing ship structure and performance to promote green, low-carbon development in the industry, thereby supporting China's "dual-carbon" goals

CHALLENGES

The ship design industry faces complex regulations, inefficient processes, and delayed delivery. Lianzhou SmartShip targets these pain points and breaks through them across compliance verification, modeling cycle, and diverse demand dimensions. With compliance assurance, intelligent efficiency improvement, and innovation empowerment at its core, the project builds a "pain point to solution" loop to improve efficiency, demonstrate data value, and drive intelligent transformation of ship design.



Industry Status and Core Pain Points

Complex Regulations and High Compliance Difficulty

The shipbuilding sector's **regulations and standards are numerous** and frequently updated, making design compliance highly challenging. Traditional search methods **struggle to pinpoint precise clauses**, and compliance checks **consume excessive time.**

Cumbersome Design Processes and Low Efficiency

Ship design **requires repeated modeling and data conversion; poor information handover and difficult collaboration** lead to long cycles. Manual 2D drawing annotation can **exceed 8 hours** per sheet; 3D modeling can take **over 60 days**, with rework rates **up to 15%.**

Insufficient Innovation Capacity to Meet Diverse Demands

Market demand for diverse and customized ship functions is rising, **with customized orders exceeding 60%.** Manually producing a design specification **takes at least 15 days, making timely responses difficult.**

Targeted Empowerment by Lianzhou SmartShip

Expert Assurance for Compliant Design

Backed by **a robust knowledge base and intelligent retrieval,** Lianzhou SmartShip integrates and continuously updates regulatory standards to enable **rapid lookup** of relevant clauses, **reducing design cycles by 25%.**

Intelligent Tools to Enhance Design Efficiency

By integrating visual recognition and 3D conversion technologies, Lianzhou SmartShip automates **2D to 3D model conversion** and **shortens modeling cycles by 50%.**

Innovation Engine to Empower Product Upgrades

Focusing on personalized market needs, Lianzhou SmartShip standardizes requirement dimensions, extracts key demand points, and **auto-generates design schemes—improving work efficiency by 90%.**

Figure 1: Industry challenges

SOLUTION AND VALUE

01 Overall Structure/Solution

Lianzhou SmartShip is supported by "compute resources + models + data" and builds a ship design digital-intelligence collaborative engine that spans R&D and design, production and manufacturing, and marketing and services. It reshapes design and service efficiency, establishes a digital-intelligence competitive moat for the industry, and leads the upgrade and transformation of ship design and manufacturing.

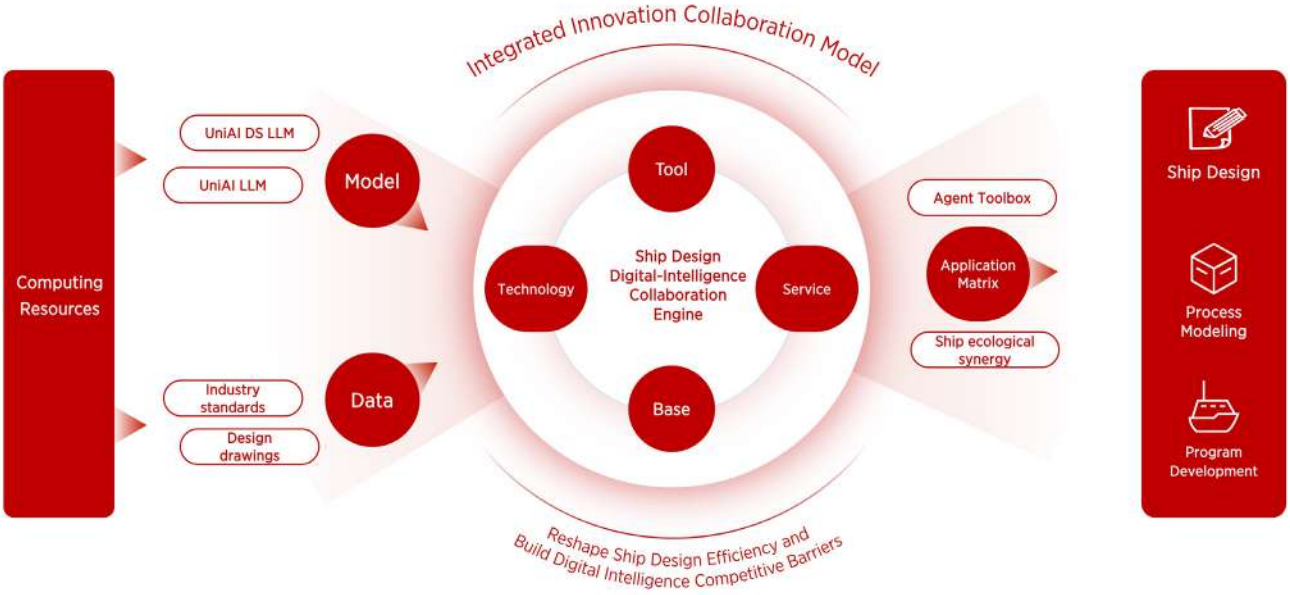

Integrated Innovation Collaboration Model
Computing Resources
UniAI DS LLM
UniAI LLM
Model
Industry standards
Design drawings
Data
Tool
Technology
Ship Design Digital-Intelligence Collaboration Engine
Service
Base
Agent Toolbox
Application Matrix
Ship ecological synergy
Ship Design
Process Modeling
Program Development
Reshape Ship Design Efficiency and Build Digital Intelligence Competitive Barriers

Figure 2: Ship design digital-intelligence collaborative engine

Product architecture

Lianzhou SmartShip adopts a five-layer product architecture: "Agent Matrix — Tool Components — Model Algorithms — Platform Base — Infrastructure," with the UniAI large model at its core. It covers the full process of regulation Q&A, modeling design, and solution assistance. Through agent collaboration, toolchain support, and algorithm engine empowerment, it reshapes the data-driven productivity of ship design.

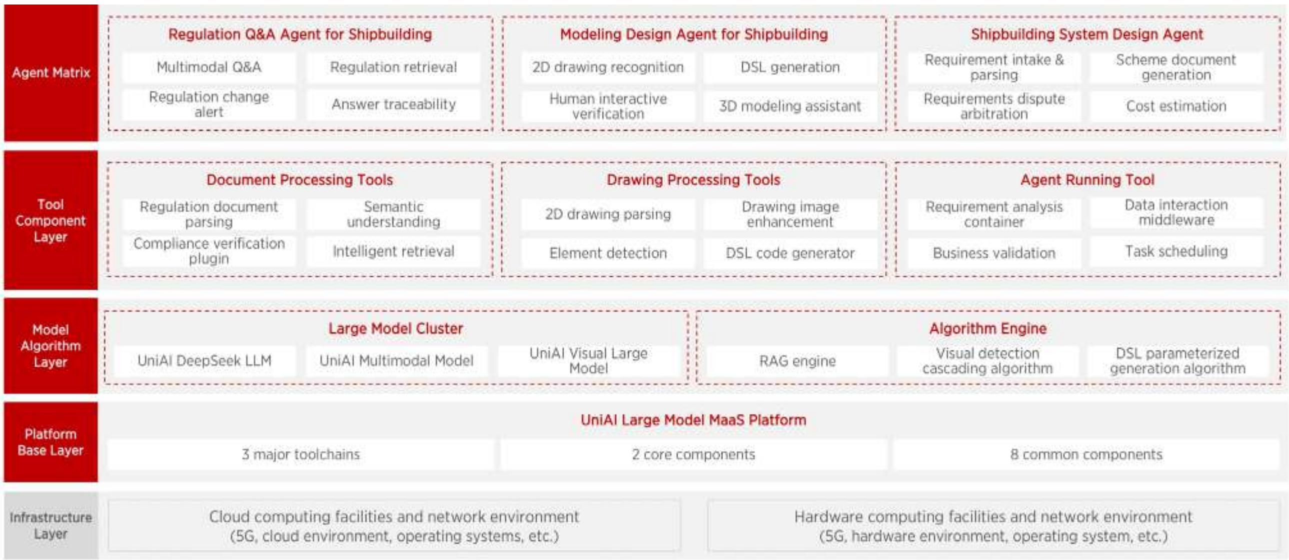

Agent Matrix
Regulation Q&A Agent for Shipbuilding
Multimodal Q&A
Regulation retrieval
Regulation change alert
Answer traceability
Modeling Design Agent for Shipbuilding
2D drawing recognition
DSL generation
Human interactive verification
3D modeling assistant
Shipbuilding System Design Agent
Requirement intake & parsing
Scheme document generation
Requirements dispute arbitration
Cost estimation
Tool Component Layer
Document Processing Tools
Regulation document parsing
Semantic understanding
Compliance verification plugin
Intelligent retrieval
Drawing Processing Tools
2D drawing parsing
Drawing image enhancement
Element detection
DSL code generator
Agent Running Tool
Requirement analysis container
Data interaction middleware
Business validation
Task scheduling
Model Algorithm Layer
Large Model Cluster
UniAI DeepSeek LLM
UniAI Multimodal Model
UniAI Visual Large Model
Algorithm Engine
RAG engine
Visual detection cascading algorithm
DSL parameterized generation algorithm
Platform Base Layer
UniAI Large Model MaaS Platform
3 major toolchains
2 core components
8 common components
Infrastructure Layer
Cloud computing facilities and network environment (5G, cloud environment, operating systems, etc.)
Hardware computing facilities and network environment (5G, hardware environment, operating system, etc.)

Figure 3: Lianzhou SmartShip product architecture

Agent technical solutions

Regulation Q&A agent for shipbuilding

To address complex questions within tens of thousands of pages of shipbuilding regulations, the project applies UniAI RAG techniques with cascaded chunking, adaptive table splitting and integration, and multi-path retrieval fusion to parse content and improve the UniAI model's recall of relevant materials.

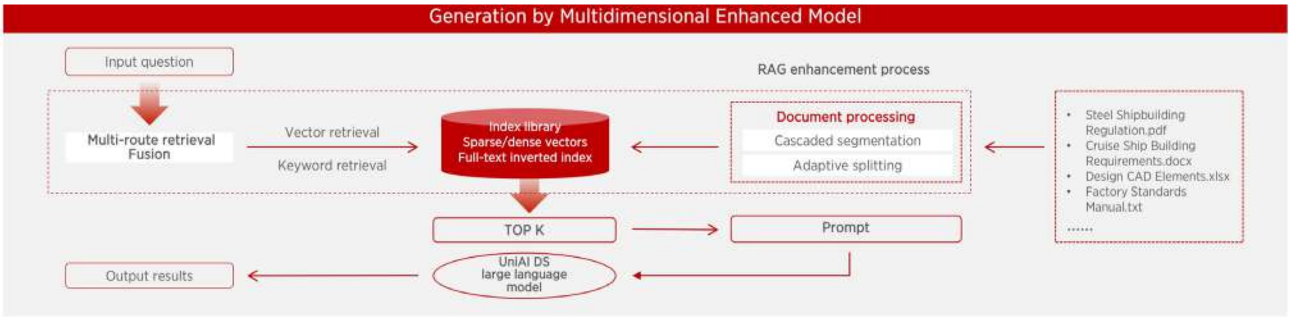

Generation by Multidimensional Enhanced Model
Input question
RAG enhancement process
Multi-route retrieval Fusion
Vector retrieval
Keyword retrieval
Index library
Sparse/dense vectors
Full-text inverted index
Document processing
Cascaded segmentation
Adaptive splitting
Steel Shipbuilding Regulation.pdf
Cruise Ship Building Requirements.docx
Design CAD Elements.xlsx
Factory Standards Manual.txt
......
TOP K
Prompt
Output results
UniAI DS large language model

Figure 4: Regulation Q&A large-model retrieval process

Modeling design agent for shipbuilding

To address the many recognizable elements in 2D ship drawings and the long 3D modeling cycle, the solution combines UniAI's multimodal element understanding, visual large-model segmentation, and visual small-model localization capabilities to automatically detect and identify the semantics and positions of key graphical elements in 2D ship design drawings, enabling designers to produce 3D models efficiently.

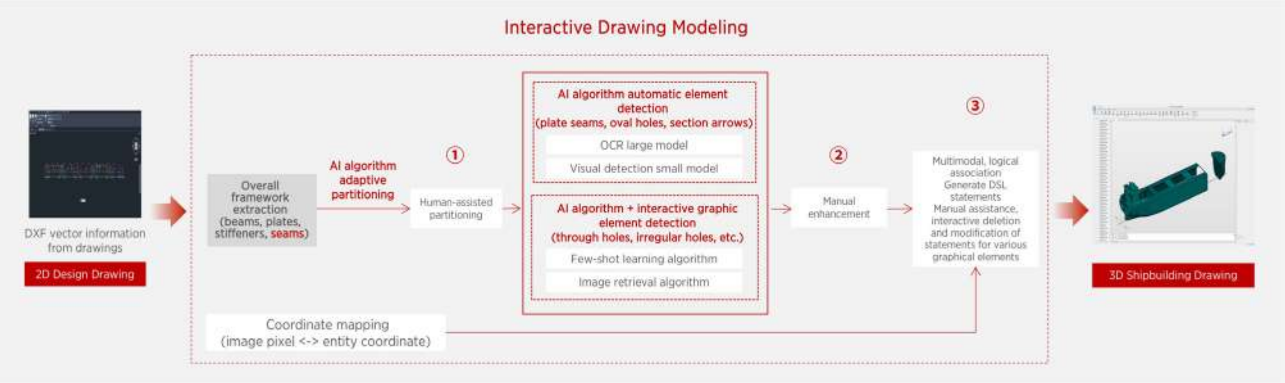

Interactive Drawing Modeling
DXF vector information from drawings
2D Design Drawing
Overall framework extraction (beams, plates, stiffeners, seams)
AI algorithm adaptive partitioning
①
Human-assisted partitioning
AI algorithm automatic element detection (plate seams, oval holes, section arrows)
OCR large model
Visual detection small model
AI algorithm + interactive graphic element detection (through holes, irregular holes, etc.)
Few-shot learning algorithm
Image retrieval algorithm
②
Manual enhancement
③
Multimodal, logical association
Generate DSL statements
Manual assistance, interactive deletion and modification of statements for various graphical elements
Coordinate mapping (image pixel <-> entity coordinate)
3D Shipbuilding Drawing

Figure 5: Modeling design workflow

Software system design agent for shipbuilding

To address the many customized requirements and long development cycles in shipboard software system development, the project combines the UniAI large language model and agent capabilities to automatically parse customer requirements and output the business elements and development specifications required for software system design, assisting more efficient system development.

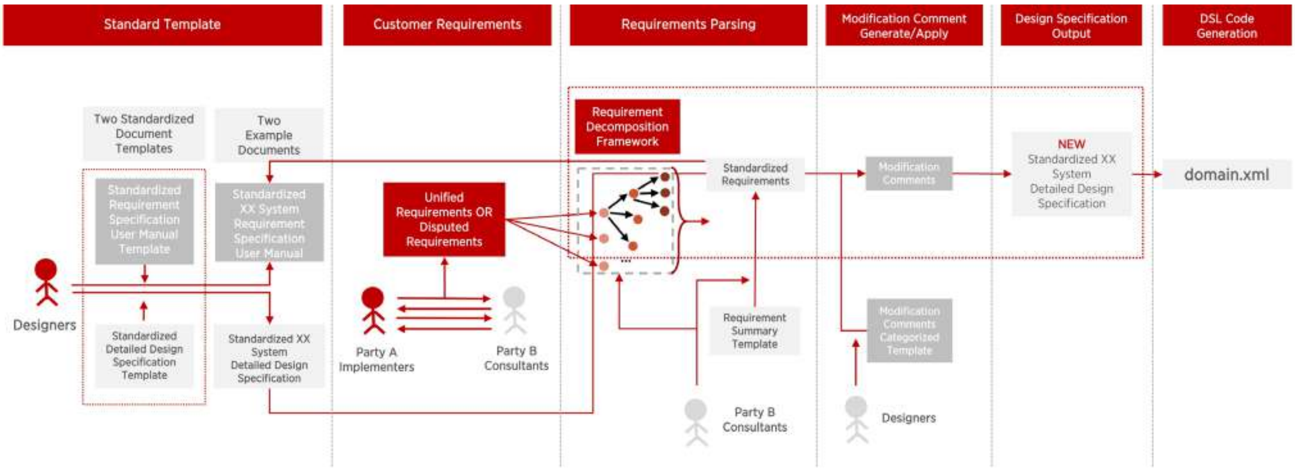

Standard Template
Customer Requirements
Requirements Parsing
Modification Comment Generate/Apply
Design Specification Output
DSL Code Generation
Two Standardized Document Templates
Two Example Documents
Standardized Requirement Specification User Manual Template
Standardized XX System Requirement Specification User Manual
Designers
Standardized Detailed Design Specification Template
Standardized XX System Detailed Design Specification
Unified Requirements OR Disputed Requirements
Party A Implementers
Party B Consultants
Requirement Decomposition Framework
Standardized Requirements
Requirement Summary Template
Party B Consultants
Modification Comments
Modification Comments Categorized Template
Designers
NEW
Standardized XX System Detailed Design Specification
domain.xml

Figure 6: Software design specification generation process

02 Application Scenarios

Regulation Q&A agent for shipbuilding

Rapid regulation retrieval	Designers can enter keywords or natural-language questions, and the large model instantly searches massive shipbuilding regulation repositories, returning relevant content from authoritative bodies including class societies, national standards, etc.
Personalized query filters	Support for custom query constraints by vessel type (e.g., tanker, container ship), construction stage (design, construction, acceptance), and other preferences to obtain precisely applicable regulations.
Answer traceability and knowledge management	Q&A results link directly to the corresponding regulation document sections; designers can jump back to source text with one click to ensure design justification. The system also supports Q&A analytics by vessel type. It can count high-frequency regulation questions and common error points.
Retrieval and answer performance	For technical difficulties across text, tables, and formulas, the product leverages Unicom UniAI RAG techniques for structural analysis, delivering industry-leading accuracy and recall.

钢质海船入级规范

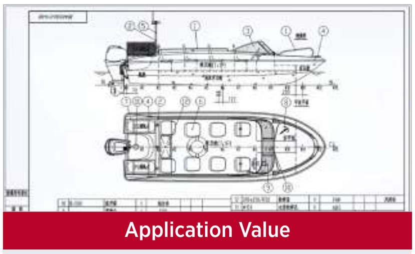
2024 第3次

Technology and Data Support

- 20,000+ shipbuilding regulation documents ingested with shared semantic understanding
- Three-level chunking strategy with semantic + positional dual indexing
- Multi-path retrieval fusion to meet cross-requirement queries

Core Features

- Multi-modal regulation knowledge Q&A: parsing of text, tables, figures, and formulas; builds professional sub-knowledge graphs
- Answer traceability and knowledge management: Q&A results link to regulation document sections; supports Q&A analytics
- End-to-end compliance check & alerting: dynamic monitoring of regulation changes and multidimensional compliance validation

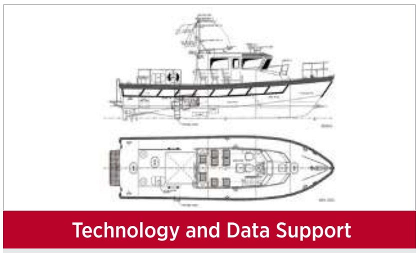


Application Value

- Q&A accuracy: **45%** -> **85%**
- Processing efficiency: **80%** ↑
- Design cycle: **25%** ↓

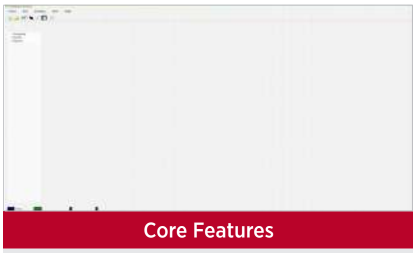
Modeling design agent for shipbuilding

Intelligent drawing reading	Rapid recognition of various 2D drawing formats, accurately parsing lines, annotations, and symbols—whether detailed mechanical parts or large assembly drawings.
Smart data conversion	Automatically converts parsed 2D drawing information into code usable by 3D modeling software.
Precision modeling assistance	During data transfer, the large model analyzes drawing information and provides optimization suggestions for 3D modeling (structure rationality, dimensional accuracy adjustments), helping designers build high-quality 3D models quickly.
Modeling performance	Enables designers to rapidly parse 2D drawings and shorten modeling cycles. Traditionally, 10 to 20 engineers require 3 to 5 months for 2D to 3D modeling; the modeling product can reduce design personnel by 40% and improve efficiency by 50%.



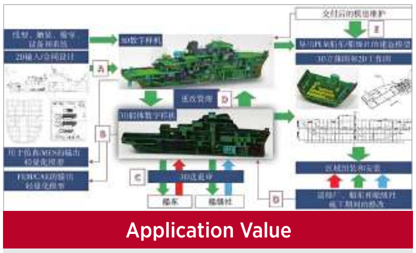
Technology and Data Support

- Injected over 100 drawings and accumulated 13+ distinct graphical element types, covering bulk carriers, tug-cable vessels, and Yangtze River cruise ships
- UniAI multimodal large model + visual large model + visual small models form a cascaded multi-level detection pipeline for robust element extraction
- Coordinate-mapping technology achieves precise conversion from image pixels to real-world entity coordinates



Core Features

- Intelligent 2D drawing parsing: adaptive drawing classification, coordinate conversion, area partitioning, and element detection
- 3D modeling assistance: automatically generate DSL modeling code from recognized elements to drive 3D software and create 3D model frameworks
- Interactive element verification: designers can manually correct DXF vector data and detected elements, enabling human-in-the-loop optimization

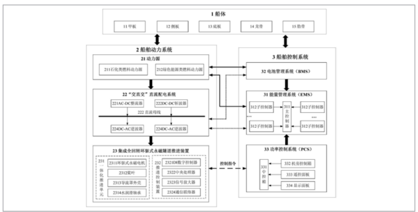


Application Value

- Graphic element recognition accuracy: **70%**
- 120 local thumbnails: **8 hours+** -> **50 minutes**
- 3D modeling cycle: **60 days+** -> **30 days**

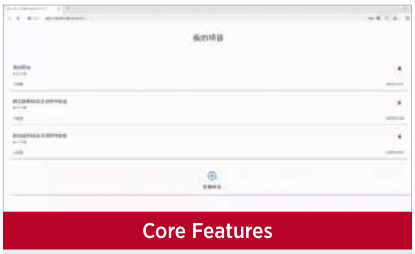
Software system design agent for shipbuilding

Intelligent scheme generation	Designers provide basic parameters (vessel type, purpose, tonnage, navigation area), and the large model quickly retrieves and matches industry best-practice cases to auto-generate complete ship design schemes that comply with industry standards.
Agile scheme generation and composition	Using China Unicom's agent orchestration, the system rapidly generates ship design documents in combined text-and-graphic formats.
System design performance	Generalizes quickly to different business system requirements; runtime efficiency is roughly 10 times that of comparable products (such as Dify).



Technology and Data Support

- Parse 20+ types of business systems including system data, collaboration management, etc.
- Build requirement-specification-solution linking graph
- Automatic linking of standard product parameters



Core Features

- In-depth requirement parsing & expansion: integrates and prioritizes multidimensional requirements, automatically decomposes ambiguous requests into detailed function points, and links to industry best practices to enrich requirement parameters
- Automatic design document generation: produces system design specifications, cost estimation reports, and full-process documentation automatically
- Requirement dispute arbitration & validation: detects conflicting requirements and invokes the regulation knowledge graph plus historical data to arbitrate, outputting a "compliance assessment + alternative solutions"

Application Value

- Requirement parsing time: **2 days** -> **0.5 hours**
- Scheme document generation: **15 days** -> **1 hour**
- Scheme compliance rate: **80%**

03 Key Innovations

① Modeling design is an industry-first technical innovation. In ship design, annotating process parameters on 2D drawings and producing 3D models usually requires 10 to 20 engineers and 3 to 5 months. China Unicom uses UniAI multimodal and visual large-model capabilities to recognize parameters and design elements on 2D drawings and output data elements required by Seagosoft's 3D modeling. Current coverage includes 2D element recognition and mapping to the modeling DSL, saving 40% of designers and boosting efficiency by 50%.

② Software design assistant is the industry's first intelligent agent for agile system development. Shipboard software development typically spans 10+ business systems and multiple departments; requirement changes slow system design. China Unicom's agent orchestration technology fuses requirements and system development knowledge to create a collaborative, document-centric agent system across scenarios, improving agile development efficiency by 50%.

04 Business Models

Target customers	Revenue models
Shipbuilding companies and central state-owned enterprises (SOEs) and small- and medium-sized enterprises (SMEs) in the industrial manufacturing sector with needs for complex regulation Q&A, drawing modeling, or agile software system development management.	• **Standardized product promotion:** Market packaged products for regulation Q&A, drawing modeling, and agile software system development to industrial manufacturing central SOEs. • **Joint R&D and operations:** Co-develop tailored solutions with enterprise customers in matching business scenarios and share operating revenue.

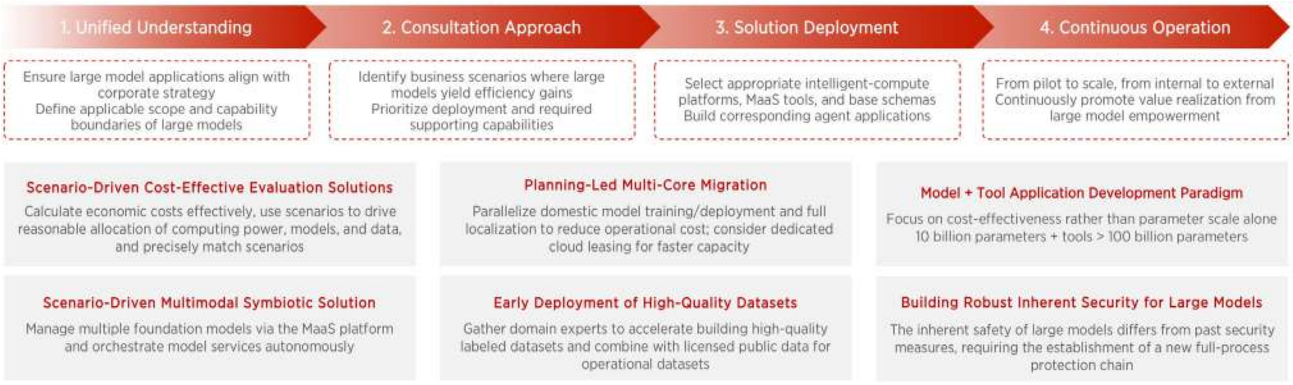
05 Core Values

Economic value	Technical value	Industrial value	Ecological and social value
Estimated labor savings per new ship design over 2 million yuan; design cycle shortened by 2 to 3 months. With three ships delivered per year, annual cost reduction is around 7 million yuan.	Builds an industry knowledge graph and a reusable agent library, laying groundwork for migration to sectors such as automotive and high-speed rail.	Helps diversify and sustain the global ship design technology supply chain.	Supports green low-carbon design—reducing aluminum and steel waste by about 5% to 8% annually—and cultivates a workforce of AI + shipbuilding hybrid talents on the order of thousands.

REFLECTION

Key Insights

Enter via focused scenarios and deepen cooperation



Ensure large model applications align with corporate strategy
Define applicable scope and capability boundaries of large models

Identify business scenarios where large models yield efficiency gains
Prioritize deployment and required supporting capabilities

Select appropriate intelligent-compute platforms, MaaS tools, and base schemas
Build corresponding agent applications

From pilot to scale, from internal to external
Continuously promote value realization from large model empowerment

Scenario-Driven Cost-Effective Evaluation Solutions
Calculate economic costs effectively, use scenarios to drive reasonable allocation of computing power, models, and data, and precisely match scenarios

Planning-Led Multi-Core Migration
Parallelize domestic model training/deployment and full localization to reduce operational cost; consider dedicated cloud leasing for faster capacity

Model + Tool Application Development Paradigm
Focus on cost-effectiveness rather than parameter scale alone
10 billion parameters + tools > 100 billion parameters

Scenario-Driven Multimodal Symbiotic Solution
Manage multiple foundation models via the MaaS platform and orchestrate model services autonomously

Early Deployment of High-Quality Datasets
Gather domain experts to accelerate building high-quality labeled datasets and combine with licensed public data for operational datasets

Building Robust Inherent Security for Large Models
The inherent safety of large models differs from past security measures, requiring the establishment of a new full-process protection chain

Figure 7: Reflections

Future Plans

Three-year roadmap

Year 1 Deploy and operate with 5 to 10 key shipbuilding customers to register a revenue of 10 million yuan, cover and accumulate in key clients and industries to iterate smart ship project agents.

Year 2 Expand maritime agents and relevant products into broader industrial sectors (automotive, high-speed rail, and other high-precision manufacturing industries), targeting a network of 20 to 30 customers and a revenue of 30 million yuan.

Year 3 Form a cross-industry platform, consolidate vertical agents, open APIs for rapid adaptation and scaling, and target another 30 million yuan revenue. Drive AI and general industrial manufacturing ecosystem growth across design, verification, manufacturing, operations, and maintenance, creating employment opportunities for tens of thousands.

AI-Powered Environmental Protection: Practice of the Xingchen Eco-Environment Governance Large Language Model

> **"Xingchen Eco-Environment Governance Large Language Model (LLM)" is an AI-based platform for ecological environment governance. It targets the environmental protection sector, combines digital twin technology with algorithm models, and delivers functions including real-time monitoring and alerts, pollution source tracking, and disaster warnings, advancing environmental protection toward greater scientific rigor, precision, and intelligence.**
>
> **Wei Bangcai**
> **China Telecom Wanwei Information Technology Co., Ltd.**
> **Party committee member, deputy general manager**

SOLUTION
PARTNERS

中国电信 CHINA TELECOM | 中电万维 CHINA TELECOM WANWEI

The Department of Ecology and Environment of Hebei Province actively responded to the new developments and requirements in ecological and environmental protection. In accordance with the requirements of the "Notice on Strengthening the Construction of Municipal and County-Level Air Command and Dispatch Platforms" [Ji Huan Ban Zi Han (2023) No. 82], it fully promoted the construction of the air command and dispatch platform, aiming to enhance the precision and efficiency of air pollution control. The platform's core functionalities include monitoring ambient air quality, supervising emission sources, managing tasks, issuing early warnings for high-concentration hotspots, dispatching enforcement actions, and conducting precise source tracing. The objective is to create unified air quality oversight and coordinated pollution source management, achieving real-time alerts for exceedances, accurate identification of polluting enterprises, and advanced features like one-click dispatch, thereby offering robust support for air pollution control. Furthermore, the platform is deeply integrated with the Xingchen Eco-Environment Governance LLM, significantly boosting its intelligent capabilities. Since the trial version of the model was deployed across cities and counties, the platform has been used extensively, logging 1,974 uses. Key users include the Baoding and Handan Ecology and Environment Bureaus, which applied the platform to essential operations like source analysis, air quality monitoring, task handling, and data alert management. The platform's efficient operation has substantially increased the effectiveness and precision of environmental supervision, transforming local air pollution control efforts.

CHALLENGES

Data Silos and Inconsistent Quality

Data in the environmental protection sector is scattered across various departments including environmental protection, meteorology, and enterprise emissions. The absence of standardized formats and varying update frequencies complicates data integration. For example, emission data from some enterprises is still manually reported, causing problems like underreporting and missed entries. The mismatch in spatial and temporal resolution between weather data and air quality monitoring data reduces the precision of source analysis. Furthermore, missing historical data or irregular annotations hinder the training performance of AI models.

Delayed Monitoring and Emergency Response

Traditional environmental monitoring depends on fixed-point stations, which have restricted coverage and are inadequate for tracking rapid changes during sudden pollution incidents. For example, short-term high-intensity pollution such as VOC leaks in industrial zones or straw burning are frequently identified only after they have spread, resulting in delayed emergency measures. Additionally, emergency response plans are typically based on human expertise without a data-driven adaptive system, making them insufficient for addressing complex pollution hazards.

Lack of Transparency and Effectiveness in Law Enforcement

Environmental violations are highly concealed. Traditional enforcement methods depend heavily on spot checks, requiring high labor costs and being prone to subjectivity. For example, there is a lack of technical tools for detecting activities such as "stealth emissions at night" or "falsification of data," resulting in lengthy investigation processes and insufficient punitive impact. Additionally, mechanisms for inter-regional enforcement coordination are immature, and delays in information exchange impede the formation of a unified regulatory effort.

Challenges in Technology Integration and Real-World Application

The integration of new technologies such as AI and digital twins with environmental protection practices is still in the exploratory stage, with a tendency to emphasize technology over application. While certain large AI models have pollution forecasting abilities, the lack of deep integration with real-world contexts such as law enforcement and emergency management limits their potential. Moreover, the low adaptability of frontline environmental staff to new technologies and complex operations constrains widespread adoption.

SOLUTION AND VALUE

01 Overall Structure/Solution

The Xingchen Eco-Environment Governance LLM is built upon an industry-specific model, which delivers core services such as air and water quality forecasting. The middle layer establishes basic functionalities, including model management, data management, data cleansing and sharing, supporting the model's training, assessment, optimization, and deployment. The top layer enables the integration into various scenarios and application services, covering multiple domains including atmospheric environment, water environment, comprehensive law enforcement, and intelligent environmental impact assessment. The platform also features scenario-specific services such as digital humans, an environmental knowledge base, and AI-assisted enforcement discretion, alongside capabilities like AI-powered search and Q&A, aiding environmental monitoring, emergency response, and pollution source tracing.

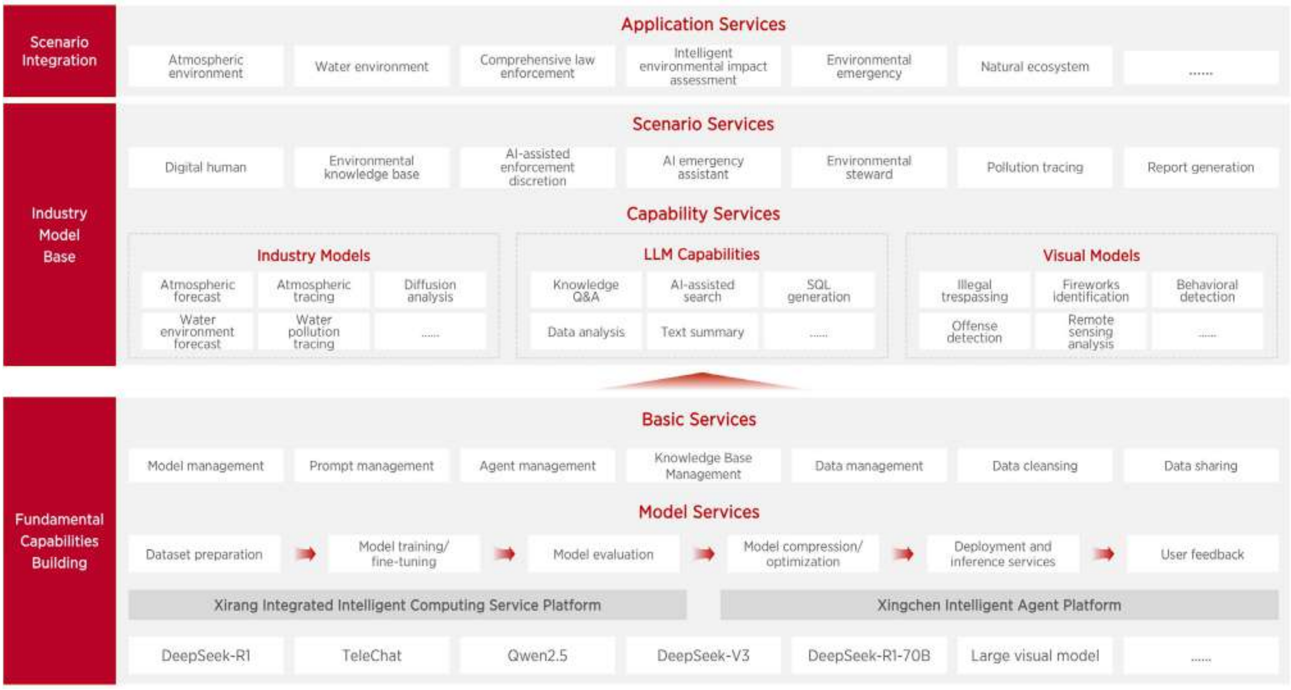
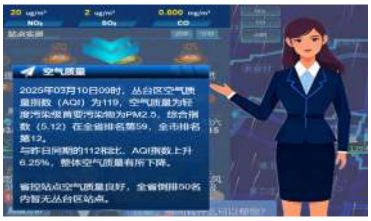

Scenario Integration
Application Services
Atmospheric environment
Water environment
Comprehensive law enforcement
Intelligent environmental impact assessment
Environmental emergency
Natural ecosystem
......
Industry Model Base
Scenario Services
Digital human
Environmental knowledge base
AI-assisted enforcement discretion
AI emergency assistant
Environmental steward
Pollution tracing
Report generation
Capability Services
Industry Models
Atmospheric forecast
Atmospheric tracing
Diffusion analysis
Water environment forecast
Water pollution tracing
......
LLM Capabilities
Knowledge Q&A
AI-assisted search
SQL generation
Data analysis
Text summary
......
Visual Models
Illegal trespassing
Fireworks identification
Behavioral detection
Offense detection
Remote sensing analysis
......
Fundamental Capabilities Building
Basic Services
Model management
Prompt management
Agent management
Knowledge Base Management
Data management
Data cleansing
Data sharing
Model Services
Dataset preparation
Model training/ fine-tuning
Model evaluation
Model compression/ optimization
Deployment and inference services
User feedback
Xirang Integrated Intelligent Computing Service Platform
Xingchen Intelligent Agent Platform
DeepSeek-R1
TeleChat
Qwen2.5
DeepSeek-V3
DeepSeek-R1-70B
Large visual model
......

Figure 1: Overall architecture

02 Application Scenarios

Digital human

Integrating DeepSeek, the digital human "Xiaoqing" is created. It incorporates advanced digital human capabilities to schedule push notifications of environmental data like air quality, high-value warnings, and task status. It can be activated by voice commands "Xiaoqing" or "Xiaoqing Assistant" to enable real-time conversations for querying monitoring data, acquiring management strategies, and executing commands, while also pushing routine reports and urgent alerts, establishing a smart environmental governance system.


20 ug/m³
NO₂
2 ug/m³
SO₂
0.800 mg/m³
CO

空气质量

2025年03月10日09时，丛台区空气质量指数（AQI）为119，空气质量为轻度污染级首要污染物为PM2.5，综合指数（5.12）在全省排名第59，全市排名第12。
与昨日同期的112相比，AQI指数上升6.25%，整体空气质量有所下降。

省控站点空气质量良好，全省倒排50名内暂无丛台区站点。

Figure 2: Digital human

Environmental knowledge base

The environmental knowledge base employs the Xingchen Eco-Environment Governance LLM to consolidate industry data, delivering thorough and up-to-date environmental information to support decision-making and technology innovation. Its agent can rapidly assess environmental quality data and use visual displays to supply accurate data for modern environmental governance.

Figure 3: Environmental knowledge base

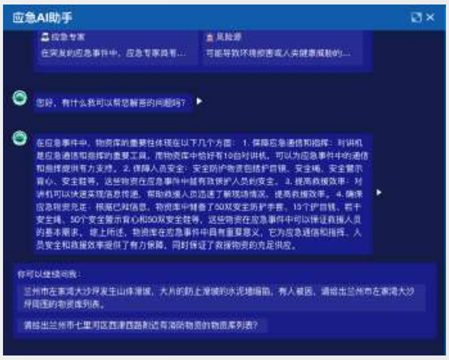
Figure 4: AI-assisted enforcement discretion

AI-assisted enforcement discretion

Intelligent AI-assisted enforcement discretion utilizes the Xingchen Eco-Environment Governance LLM to perform real-time analysis of whether a current activity breaches environmental rules and regulations, supplying detailed information. The AI analysis can precisely determine the legal grounds for the violation and provide a comprehensive explanation, as well as quantify the penalty level. This offers law enforcement officers a scientific and objective foundation, leading to fairer and more effective environmental enforcement, greatly improving its precision and efficiency, and providing robust support for protecting the environment and maintaining legal order.

Environmental emergency assistant

Utilizing the robust data processing and analytical power of the Xingchen Eco-Environment Governance LLM, a holistic and smart emergency AI assistant is built. It consolidates diverse emergency data including emergency supply inventories, experts, response teams, historical cases, contingency plans, environmental risk sources, and hazardous chemicals, substantially boosting the efficiency and outcomes of environmental emergency disposal. This aids the shift from traditional sandtable command to intelligent command, enhancing risk prevention during peacetime and guaranteeing swift, precise command and dispatch during emergencies, offering solid support for environmental safety and social stability.



应急AI助手

应急专家
在突发的应急事件中，应急专家具有…

风险源
可能导致环境损害或人类健康威胁的…

您好，有什么我可以帮您解答的问题吗？

在应急事件中，物资库的重要性体现在以下几个方面：1. 保障应急通信和指挥：对讲机是应急通信和指挥的重要工具，而物资库中恰好有10台对讲机，可以为应急事件中的通信和指挥提供有力支持。2. 保障人员安全：安全防护物资包括护目镜、安全绳、安全警示背心、安全鞋等，这些物资在应急事件中能有效保护人员的安全。3. 提高救援效率：对讲机可以快速实现信息传递，帮助救援人员迅速了解现场情况，提高救援效率。4. 确保应急物资充足：根据已知信息，物资库中储备了50双安全防护手套，10个护目镜，若干安全绳，50个安全警示背心和50双安全鞋等，这些物资在应急事件中可以保证救援人员的基本需求。综上所述，物资库在应急事件中具有重要意义，它为应急通信和指挥、人员安全和救援效率提供了有力保障，同时保证了救援物资的充足供应。

你可以继续问我：

兰州市在寨湾大沙坪发生山体滑坡，大片的防上滑坡的水泥墙垮塌，有人被困，请给出兰州市在寨湾大沙坪周围的物资库列表。

请给出兰州市七里河区西津西路附近有消防物资的物资库列表？

Figure 5: Environmental emergency assistant

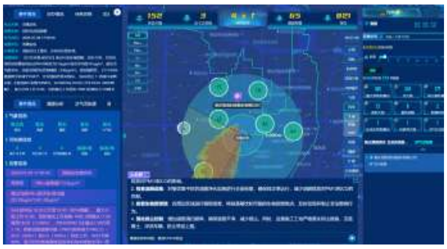


Figure 6: Source analysis

Source analysis

By embedding DeepSeek into the foundation of the environmental protection cloud product, a "data collection - multi-modal integration - algorithmic inference - accurate source identification" closed loop is formed. The Xingchen Eco-Environment Governance LLM, employing its mixture-of-experts (MoE) architecture and long chain-of-thought reasoning capacity, consolidates environmental monitoring data, industrial discharge data, and weather data, enabling parallel computing for multiple specific models. Coupled with high-precision computation and causal reasoning capabilities, and incorporating the specialized STILT traceability model, it accomplishes precise tracing of environmental pollutants.

Report generation

Based on natural language processing and machine learning, the Xingchen Eco-Environment Governance LLM and the report generation feature are developed. The feature gathers real-time multi-source data from environmental monitoring devices and company emission data. It automatically produces formatted, content-rich environmental reports derived from analysis results. The report templates and key content can be flexibly adjusted according to different user needs, greatly improving the efficiency and quality of environmental reports, providing strong support for environmental protection decision-making.

Figure 7: Report generation

03 Key Innovations

Multi-modal data integration and long chain-of-thought reasoning

Leveraging the MoE architecture of the Xingchen Eco-Environment Governance LLM, it consolidates diverse data sources including environmental monitoring, industrial emission, and weather information, forming a closed loop from data collection and multi-modal integration to algorithmic inference and accurate source identification. In pollution source tracing scenarios, by combining the STILT model with causal inference, it enables dynamic simulation of pollutant transmission routes, boosting tracing accuracy to 90%.

Intelligent interaction system combining digital human and knowledge base

The digital human "Xiaoqing" is developed, capable of voice activation and real-time conversation. It automatically delivers alerts on air quality, task status updates, and connects to the environmental knowledge base for decision support. In periods of severe air pollution, Xiaoqing can recommend adjustments to emergency plans, helping command centers make faster decisions.

Quantitative and transparent AI-assisted enforcement discretion

Using the Xingchen Eco-Environment Governance LLM coupled with knowledge base technology, it assesses violation attributes in real-time, automatically identifies applicable legal clauses, quantifies the grounds for penalties, and generates legal documentation. For instances of "exceedance emissions," the system can dynamically calculate fines by considering factors like emission concentration and duration, minimizing manual intervention.

04 Business Models

Project-based model

A one-time fee is charged for localized project deployment. Tailored development can be conducted according to individual customer requirements, with costs negotiated case-by-case. Offerings include digital human, environmental knowledge base, AI-assisted enforcement discretion, environmental emergency assistant, source analysis, report generation, and personalized requirements.

SaaS model

Standardized products are charged with an annual platform service fee. This covers platform data integration and customer account setup, but does not include custom development for specific needs. Services include digital human, environmental knowledge base, AI-assisted enforcement discretion, environmental emergency assistant, source analysis, and report generation.

05 Core Values

Quantitative results

- Business processing efficiency has increased by **65%** (in scenarios involving environmental knowledge Q&A).
- Emergency plan adoption rate has reached **85%,** with response time reduced by **50%**.
- Pollution source tracing accuracy has reached **90%**, enabling foundational solutions to pollution issues.
- Law enforcement transparency and efficiency have increased by **40%**, and the penalty compliance rate has improved by **30%**.

Qualitative impact

- Drives the shift in environmental regulation from "experience-driven" to "data-driven", upgrading scientific decision-making.
- Lowers the operational barrier for field operators by utilizing tools such as digital humans and knowledge bases, promoting broader technology adoption.
- Establishes a smart, end-to-end framework covering "monitoring, alerting, traceability, enforcement, and emergency response," setting a benchmark for the industry.

REFLECTION

Key Insights

The integration of technology should be tightly focused on operational difficulties, driven by critical requirements including pollution source tracing and emergency response. Utilizing specific technologies such as the MoE architecture helps resolve issues like multi-source data integration, preventing piling up of technical parameters. Implementing solutions within specific scenarios is crucial for translating AI value; AI functions are decomposed into concrete components such as AI-assisted enforcement discretion and report generation. The voice-interactive digital human "Xiaoqing" simplifies usability, enabling field staff to quickly adapt without formal training. Collaborative ecosystems help speed up technology value realization. Cooperating with partners like DeepSeek improves model capabilities, and jointly building data-sharing mechanisms with local governments effectively tackles "data isolation," fostering deep collaboration between technology and practical applications.

Future Plans

By the end of 2025, we will complete the lightweight optimization of the model to enable deployment on edge devices, and investigate uses for multimodal LLMs in ecosystem restoration contexts, improving real-time response and application scope. During 2026, we will extend the proven approach to water environment and soil pollution prevention, establishing a unified sky-air-ground monitoring system to broaden operational boundaries. We plan to launch a "free basic platform + paid premium services" offering, supplying tailored emission alert services to small and medium-sized enterprises, with the goal of raising the SaaS revenue share to 40% by 2026. Also, we will partner with industry associations to release the "White Paper on AI Applications in Environmental Protection," advancing the standardization of technical norms and data interfaces to strengthen the leading market position.

TeleNavi Xingchen Wendao Large Language Model Supports Intelligent Mobility for Seniors

> **With the development of the Internet, seniors face challenges impacted by the digital divide, struggling with services such as map navigation and online ride-hailing. Traditional phone-based ride-hailing suffers from problems like order grabbing and cancelation because of imprecise location. China Telecom has pioneered the world's first voice-only positioning technology, integrating Open Gateway API with large language models (LLMs) to launch "TeleNavi Wendao", which offers precise and prompt ride-hailing for older adults, helping them overcome the digital divide and fostering the silver economy.**

Yueyang Yin
China Telecom Group
Senior AI Expert

SOLUTION PARTNERS 中国电信 CHINA TELECOM

Aiming to reduce the barriers to digital mobility for older adults, China Telecom Shandong originally conceived and independently developed the world's first voice-only positioning technology: "Key Interactive Voice-Based Map-Free Precision Positioning and Navigation Technology and Age-Friendly Application". This project combines the operator's cloud and network strengths with AI LLMs. Leveraging Open Gateway API and the proprietary TeleChat Xingchen LLM, it trained the world's first voice positioning and navigation LLM, "Wendao". It breaks through the reliance on traditional positioning technologies, achieving for the first time pure voice intelligent positioning and navigation, an industry first that fills a market gap. It achieves positioning accuracy within 5 meters, even without GPS, data connectivity, maps, or smart devices.

This technology effectively solves the problem of the 95128 voice ride-hailing hotline being unable to locate users. It offers seniors precise and prompt ride-hailing services and intelligent, voice-only navigation. This lets those left behind in the smart age experience and utilize AI, enjoy technological benefits, bridge the digital divide, remain linked to the intelligent era, and overcome mobility challenges in the digital age. It exemplifies AI for social good and a shift from virtual concepts to tangible applications, fostering senior mobility—encouraging local trips and longer journeys—which drives lifestyle evolution and consumption upgrades, aiding the silver economy. This demonstrates the corporate social responsibility of China Telecom, a Chinese central state-owned enterprise, and fulfills its mission to "enable customers to enjoy the new information lifestyle."

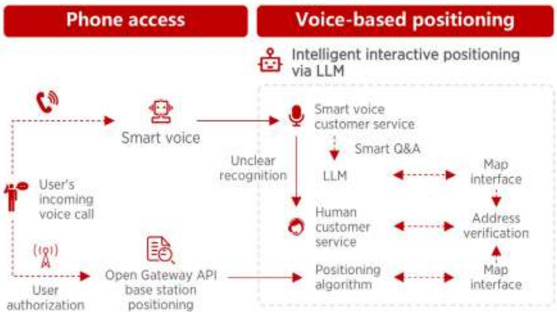


Figure 1: System flow chart 1

CHALLENGES

As global aging accelerates, the elderly population aged 60 and above exceeds 20% globally by 2024, with many countries facing aging challenges. Data shows that merely 24.1% of China's seniors use smartphones for daily assistance, 74% still flag down taxis on the roadside, and 25% travel under one kilometer daily, becoming disadvantaged in mobility. Phone-based ride-hailing faces promotion challenges due to inefficient, costly manual operations and elderly users' dialect diversity and poor direction sense, preventing precise location tracking. This leads to disorderly order grabbing and cancellations by drivers. Cross-border ride-hailing also has drawbacks. With Uber absent in China and DiDi unavailable outside China, travelers struggle to book rides. They face a series of challenges, including app downloads, registration, setup, and language issues.

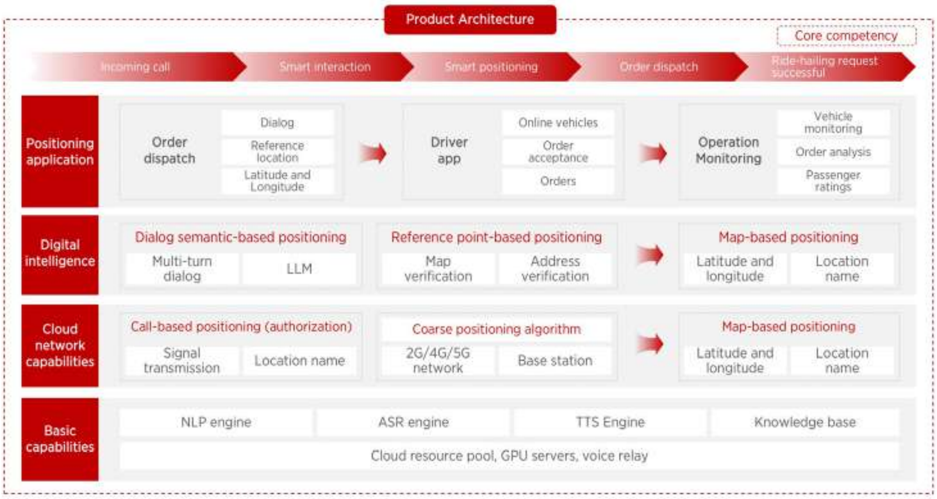
No senior-friendly features	Inaccurate location	Delayed response	Lack of operational insight
Seniors do not know how to use ride-hailing apps	Drivers often receive incorrect or imprecise passenger locations	Long order acceptance time and high cancelation rate	Taxi companies lack operational data, making it difficult to improve and refine their operations

Figure 2: Issues of 95128 phone-based ride-hailing service

In summary, current positioning technology relies too heavily on satellites (GPS, BeiDou), 4G/5G networks, smart devices, and e-maps, failing to enable voice-only positioning on 2G networks. Specialized cases like elder-friendly and cross-border mobility, emergency response, and public emergency services (such as 110 and 120 emergency services) lack adequate tools, driving demand for voice positioning solutions.

SOLUTION AND VALUE

01 Overall Structure/Solution


Product Architecture
Core competency
Incoming call
Smart interaction
Smart positioning
Order dispatch
Ride-hailing request successful
Positioning application
Order dispatch
Dialog
Reference location
Latitude and Longitude
Driver app
Online vehicles
Order acceptance
Orders
Operation Monitoring
Vehicle monitoring
Order analysis
Passenger ratings
Digital intelligence
Dialog semantic-based positioning
Multi-turn dialog
LLM
Reference point-based positioning
Map verification
Address verification
Map-based positioning
Latitude and longitude
Location name
Cloud network capabilities
Call-based positioning (authorization)
Signal transmission
Location name
Coarse positioning algorithm
2G/4G/5G network
Base station
Map-based positioning
Latitude and longitude
Location name
Basic capabilities
NLP engine
ASR engine
TTS Engine
Knowledge base
Cloud resource pool, GPU servers, voice relay

The system

- an intelligent voice customer service platform
- a positioning engine agent platform (including a voice semantic LLM)
- a ride-hailing operation and order platform
- a driver order acceptance app
- ...

Figure 3: System architecture

02 Application Scenarios

Phone-based ride-hailing for seniors	While Internet advances bring great convenience to younger people, many seniors find smartphones unfamiliar. Routine activities such as using map apps or booking rides online turn into mobility barriers, creating a digital divide. Older adults remain disadvantaged in securing transportation, frequently failing to get a taxi after long waits on the road. Existing phone-based ride-hailing systems fail to pinpoint user locations accurately, causing problems like drivers disorderly grabbing or canceling orders, which impedes service usage and expansion.
Cross-border travel ride-hailing	In addition to the ride-hailing for seniors, ride-hailing in other countries or regions also presents many inconveniences. With continuous improvements in living standards and ongoing favorable immigration policies, increasing numbers of Chinese citizens are traveling abroad and foreigners are entering China. However, due to Uber's absence in China and DiDi's lack of international operations, both inbound and outbound tourists encounter difficulties booking rides. The Wendao LLM can automatically recognize the user's language and switch to converse in that language. It eliminates the need to download any app or go through tedious registration. A simple phone call can bridge communication gaps across different languages, facilitating fast and easy ride booking for international travelers.
Voice navigation	Existing navigation applications are primarily designed for younger people, offering very few aging-adapted operations. Seniors typically have a weak sense of direction, making it hard to pinpoint their location in unfamiliar settings. Existing navigation is also prone to location deviations from GPS drift, easily causing older users to get stuck walking back and forth.
Countries, regions, and remote areas with poor network coverage	In some countries, regions, and remote areas with poor 4G/5G network coverage, GPS positioning information cannot be uploaded in time, leading to inaccurate or failed positioning. Wendao functions perfectly over just a 2G network, effectively resolving positioning and navigation failures stemming from network coverage issues.

03 Key Innovations

Dialect speech recognition

Support 16 cities in Shandong and over 40 mainstream dialects in China
Powered by powerful error correction algorithms to improve speech-to-text error tolerance

Semantics-based positioning without maps

Achieve 5-meter level positioning through landmark description (such as "on the opposite side of the KFC")
Combining base station signal and semantic and inference engines, the system works without network

Elder-friendly multi-turn interaction

Enhanced useful information guidance and reduced inquiry rounds
Two-way calling and three-party calls: voice and text notifications throughout order placement, order acceptance, and arrival

Smart order dispatching

Geofencing: Select optimal nearby drivers and prevent malicious order grabbing
Robust identity verification: Ensure peace of mind when entering the vehicle and guarantee safe arrival

Access

Connect to existing platforms via APIs without affecting their operation
Provide ride-hailing phone services and an app for drivers

End-to-end operation system

Comprehensive management console and operation platform
Driver registration management, customer rating system, operation dashboard

World's first voice-only positioning technology

China Telecom originally conceived and independently developed "Key Interactive Voice-Based Map-Free Precision Positioning and Navigation Technology and Age-Friendly Application". This project combines the operator's cloud and network strengths with AI LLMs. Leveraging Open Gateway API and the proprietary TeleChat Xingchen LLM, it trained the world's first voice-only positioning and navigation LLM, "Wendao". It breaks through the reliance on traditional positioning technologies, achieving for the first time pure voice intelligent positioning and navigation, an industry first that fills a market gap. It achieves positioning accuracy within 5 meters, even without GPS, data connectivity, maps, or smart devices. The Open Gateway API's Location Retrieval function first obtains the user's general area; then the LLM interacts with the user to achieve rapid, precise positioning, significantly cutting down interaction complexity and computational inference needs, thus improving user perception.

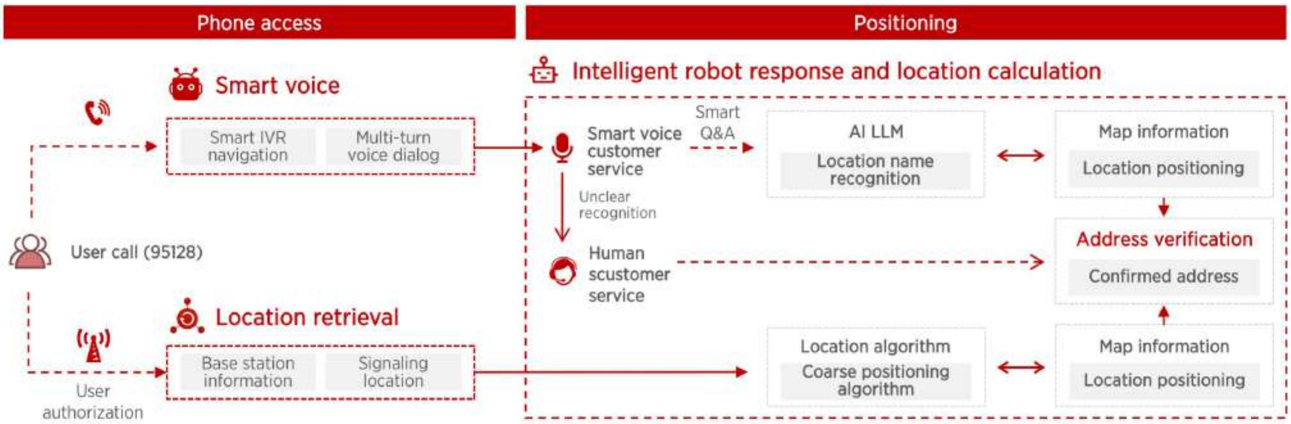
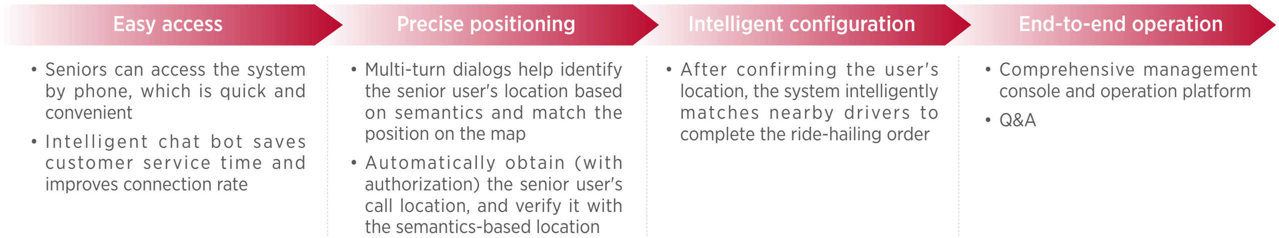

Phone access
Positioning
Smart voice
Smart IVR navigation
Multi-turn voice dialog
User call (95128)
Location retrieval
User authorization
Base station information
Signaling location
Intelligent robot response and location calculation
Smart voice customer service
Smart Q&A
Unclear recognition
Human scustomer service
AI LLM
Location name recognition
Map information
Location positioning
Address verification
Confirmed address
Location algorithm
Coarse positioning algorithm
Map information
Location positioning

Figure 4: System flow chart 2

One-stop ride-hailing service

Using Open Gateway API's Geofencing and One Time Password SMS features, the platform dispatches orders to drivers within a 3-5 km radius of the user. Once a driver accepts, the passenger receives an SMS with the car's plate number, model, and a verification code (One Time Password SMS). Triggered by a geofence when the driver is within 1 km, the platform makes a simultaneous VoLTE call to the driver and passenger, establishing communication for easy identification. Passengers verify their identity upon boarding using the code or the last four digits of their phone number, completing the ride-hailing service process.


Easy access
Precise positioning
Intelligent configuration
End-to-end operation

- Seniors can access the system by phone, which is quick and convenient
- Intelligent chat bot saves customer service time and improves connection rate

- Multi-turn dialogs help identify the senior user's location based on semantics and match the position on the map
- Automatically obtain (with authorization) the senior user's call location, and verify it with the semantics-based location

- After confirming the user's location, the system intelligently matches nearby drivers to complete the ride-hailing order

- Comprehensive management console and operation platform
- Q&A

Figure 5: One-stop ride-hailing process

Accuracy improvement for positioning and downstream tasks

Positioning accuracy is determined by the accuracy of a series of downstream tasks, including:

- Automatic speech recognition (ASR)
- Intent recognition
- Extracting location keywords
- Multi-turn conversation
- Location semantic similarity
- Question generation

Training data is gathered from routine operation data (call recordings), enabling iterative model retraining and continuous enhancement of positioning precision.

Scenario classification and template generation

The ride-hailing positioning scenarios can be classified into three categories: urban, suburban, and campus/compound areas, with further subdivisions to suit various interaction contexts. For each scenario type, we employ manually crafted templates supplemented by LLM-generated Q&A patterns. Tailored interaction templates are created considering the environment (using coarse location data and pre-fetched map information), enabling contextual and personalized user engagement. The goal is to guide the user to identify nearby references with minimal conversational turns and simple questions, thereby lowering interaction complexity and computational demands for the model.

Inverted Q&A and research on generalization of language habits

Standard LLMs typically follow a pattern where the user queries and the model responds. Our model does the opposite: the model queries and the user responds. While the questions are quite similar, the answers exhibit poor similarity due to individuals' unique expressive habits, phrasing preferences, and perspectives, leading to vastly different replies from different users. Manually deriving consistent patterns from these responses is challenging. Through iterative model training using continually collected daily data, the model progressively adapts to the diverse language customs and expressive traits of a growing user base. The development of novel data annotation techniques and formats enhances the model's generalization capability for users' language habits.

Enhancing ASR accuracy for dialects

ASR serves as the gateway to the interaction system. Dialects are prevalent among older adults, and China's dialects are immensely diverse and evolve quickly, as the saying goes, "dialects shift every three miles, accents change every ten." Consequently, constant updates and expansion of the speech corpus are essential to incorporate more dialectal variations, which trains the ASR model to recognize dialects with higher accuracy.

04 Business Models

Leveraging AI LLMs and cloud-network infrastructure, the TeleNavi Wendao model provides business-to-business (B2B) services for taxi drivers, opening up new service channels and order sources. It generates revenue by taking a commission from each order, while also stimulating growth in fundamental services such as telecom subscriptions, AI presentation services, and digital transformation solutions for transportation. On the consumer side, it enables ride booking for seniors, promoting mobility, improving living standards and consumption, thereby supporting the silver economy.

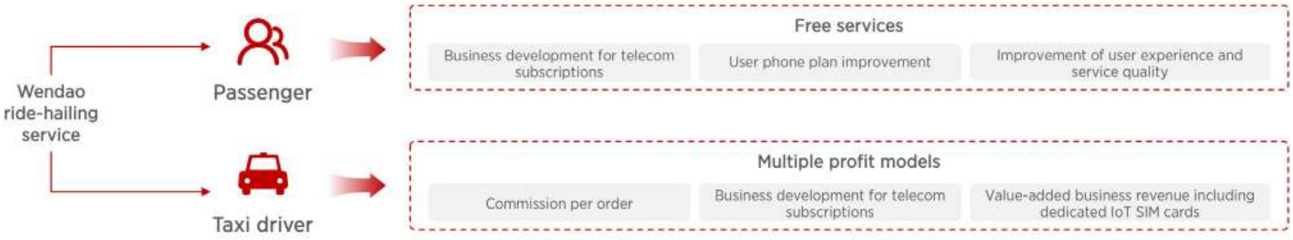

Wendao ride-hailing service
Passenger
Taxi driver
Free services
Business development for telecom subscriptions
User phone plan improvement
Improvement of user experience and service quality
Multiple profit models
Commission per order
Business development for telecom subscriptions
Value-added business revenue including dedicated IoT SIM cards

Figure 6: Business model

05 Core Values

Innovative and independently developed

Positioning and navigation LLM

Original end-to-end technology, with full control over critical processes

Simple integration

No GPS/data traffic restrictions, compatible with phones for seniors

Senior-friendly application

Caring for senior mobility, supporting the silver economy - MIIT's "age-friendly modification of electronic products"

Technology for everyone

LLMs empower seniors to enjoy technological benefits in the intelligent era

The Wendao LLM achieves accurate voice-based positioning without relying on GPS, digital maps, data connectivity, or smart devices, facilitating easy access to intelligent services for users. It fosters technological inclusion and social equity, enabling seniors who are increasingly left behind in the smart age to experience and utilize AI, enjoy technological benefits, bridge the digital divide, remain linked to the intelligent era, and overcome mobility challenges in the digital age. It exemplifies AI for social good and a shift from virtual concepts to tangible applications. This enables older adults to enjoy mobility services, improving their quality of life and consumption levels, thus supporting the development of the silver economy.

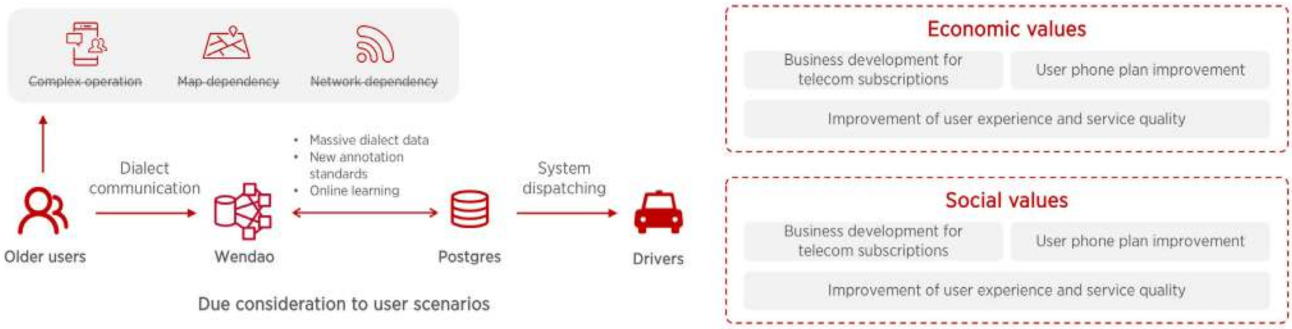

~~Complex operation~~
~~Map dependency~~
~~Network dependency~~
Older users
Dialect communication
Wendao
• Massive dialect data
• New annotation standards
• Online learning
Postgres
System dispatching
Drivers
Due consideration to user scenarios
Economic values
Business development for telecom subscriptions
User phone plan improvement
Improvement of user experience and service quality
Social values
Business development for telecom subscriptions
User phone plan improvement
Improvement of user experience and service quality

Figure 7: Economic and social values

This technology and project are highly replicable and scalable. They suit ride-hailing needs for seniors worldwide and are also applicable in developing countries and regions with limited online ride-hailing/navigation services and low smartphone coverage, offering extensive application potential.

This project has been piloted for commercial use in Rizhao and Dezhou in Shandong, and Shangzhi in Heilongjiang, saving a significant amount of customer service costs, greatly improving ride-hailing operation efficiency, and delivering good social and economic benefits. It has received consistent praise from transportation authorities, taxi companies, as well as drivers and passengers. In a single third- or fourth-tier city, daily orders range from 5,000 to 10,000, projecting annual revenues of 1-3 million yuan. Nearly 1.5 million users have been served, with expansion currently ongoing in Shandong and other provinces.

REFLECTION

Key Insights

Leveraging Open Gateway API and China Telecom's proprietary Telechat Xingchen LLM, the project effectively combined the operator's specific cloud-network strengths with AI LLMs, creating a unique competitive advantage. By utilizing internal operator assets such as coarse location services and geofencing, it tackled technical problems unsolvable by Internet or AI development companies, delivering unique value to users.

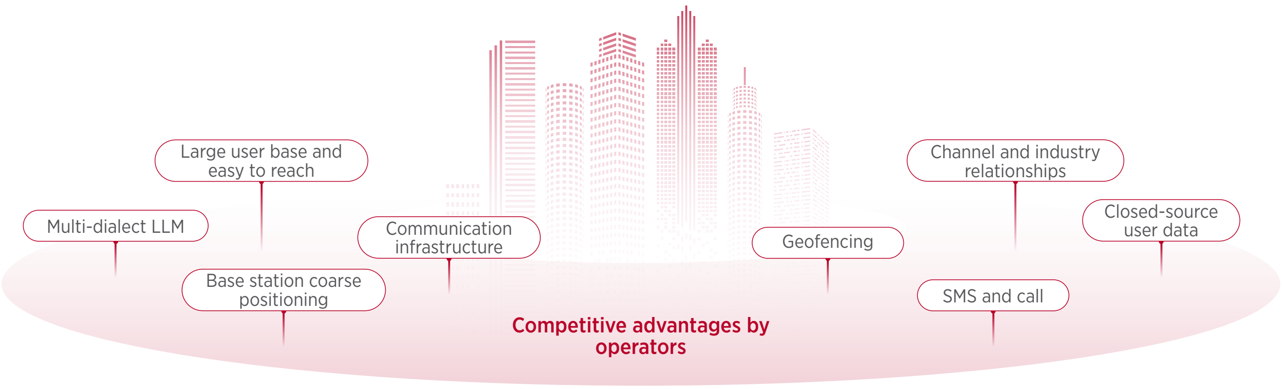

Large user base and easy to reach
Channel and industry relationships
Multi-dialect LLM
Communication infrastructure
Geofencing
Closed-source user data
Base station coarse positioning
SMS and call
Competitive advantages by operators

Figure 8: Competitive advantages

This project assists the telecom operator in transitioning from a traditional utility to a tech-oriented service enterprise amidst the AI surge, allowing customers to enjoy the new, intelligent information lifestyle. Having experienced what was called a "lost decade" during the Internet era, telecom operators need to grasp the AI opportunity, stay at the technological forefront, and integrate AI with cloud-network capabilities to offer smart information services, thereby winning the strategic shift.

Future Plans

Market expansion in China

We plan to expand the ride-hailing business across China. The target is to complete the initial deployment and operational rollout in 100 additional key prefecture-level cities by the end of 2025. This aims to cover over 1,000 taxi companies, enroll more than 100,000 drivers, and serve over 10 million users.

Application scenario expansion

The potential of "Wendao" reaches well beyond ride-hailing and navigation, offering extensive application value and replicability. Subsequent development will focus on extending its use to other areas, including community-based elderly care, emergency rescue operations, and public emergency services (such as 110 and 120 emergency services).

Emergency command and rescue management

The system can be used in areas with poor network coverage, GPS drifts, and complex urban environments. It enables location reporting via voice interaction for 110 (police) and 120 (emergency medical services). Rescue command centers can then obtain real-time positioning data.

Public services and community elderly care

It solves the travel difficulties older adults face due to inability to use navigation applications by delivering automatic route planning and voice-based navigation.

Remote areas with poor 4G/5G coverage

It provides positioning, navigation, and ride-hailing services in remote areas with no or poor 4G/5G connectivity, weak Internet signals, and developing countries/regions with limited online services.

Global expansion

Recognized as a GSMA Global Case Study benchmark and a China Telecom Group benchmark case, the project was also a GLOMO Award shortlist finalist and featured in the MIIT BRICS Outstanding AI Case Exhibition. It received invitations to showcase at the 2024 World Internet Conference (Wuzhen), 2025 TM Forum DTW (Copenhagen), 2025 ITU AI for Good Summit (Geneva), 2025 MWC Shanghai, and 2025 MWC Barcelona. Partnerships with international telecom operators are underway for deployment outside China.

International recognition and value

TeleNavi 问道

2024 World Internet Conference

Winners of second pri
¥30000

2024 MWC Hackathon Runner-up

2025 MWC Barcelona

TeleNavi
TM Forum - CHINA TE
NaaS HACKATHON
KUNLUN API + LLM

2025 TM Forum DTW

Total number of users served: 1.5 million

GSMA Global Case Study Benchmark

2025 GLOMO Award Shortlist

Continuous training and optimization

We will continuously collect user interaction data, systematically categorizing, cleansing, and selecting daily inputs for data engineering and annotation to facilitate the LLM's ongoing learning. Repeated iterative training will keep the Wendao model learning online, steadily boosting performance on downstream tasks and the efficacy of question template generation, enabling support for additional scenarios and varied linguistic patterns. Enhancing ASR coverage to include more dialects will cater to users across different geographical areas.

Large Model-Empowered Human Resources and Social Security Inquiry Assistant

> **Traditional 12333 hotline faces challenges in efficiently addressing the surging inquiry demands regarding human resources and social security policies. China Unicom has proactively integrated its optimized and upgraded cross-domain multi-source knowledge Q&A Large Language Models with a professional knowledge repository to develop an artificial intelligence inquiry assistant which capable of providing second-level responses. This case stands as an exemplary application of AI large models in vertical industries.**
>
> **Song Yulun**
> **China Unicom Data Intelligence Co., Ltd.**
> **Deputy General Manager**

SOLUTION PARTNERS

Facing the challenge of a surge in public inquiries about human resources and social security services and the overload of traditional hotlines, and in response to the national "Digital Government" strategy, China Unicom has launched a new-generation AI assistant for human resources and social security services, based on LLMs (Large Language Models) technology. This assistant aims to alleviate the burden on management departments by streamlining inquiries and unifying policy dissemination, while providing citizens and businesses with authoritative policy services that are "available around the clock, responsive in seconds, and accurate and easy to understand."

China Unicom's AI assistant adopts a "RAG + SFT" architecture, deeply integrating massive, multi-source official policy knowledge bases. This fundamentally overcomes the challenges of content "hallucination" and outdated knowledge in general LLMs, ensuring accurate responses and traceable sources. The system supports complex calculations, such as pension estimates, and multi-turn conversations. Through a collaborative "AI assistant + human expert" model, it achieves complementary advantages by providing 24/7 standardized services and in-depth support for complex issues. The solution supports private deployment, fully ensuring data security and controllability.

Implementation results are outstanding

In terms of service efficiency, a single node can independently handle over 1,200 questions per day, serving approximately 600 inquirers, equivalent to workload of 5-8 full-time staff.

From the public's perspective, the 24/7 AI assistant provides a "zero-waiting" experience for policy consultations, significantly boosting user satisfaction.

This practice, with its technical precision and model innovation, strongly promotes the digital transformation of human resources and social security services, successfully implementing the core concept of "let data run more, let people run less," and setting an important benchmark for the deep application of LLMs in the digital government field.

CHALLENGES

With the development of economics and society, human resources and social security policies have become increasingly complex, and the public's demand for refined and personalized policy services has experienced explosive growth. The hotline services across various regions are universally under pressure, and their "tidal effect" exacerbates the supply-demand contradiction. The service timeliness, cost, and coverage of traditional manual models can no longer meet the public's urgent expectation for "instant response and accurate answers," making intelligent transformation of human resources and social security services imperative.

However, on the path to intelligence, technological evolution faces dual challenges. On one hand, traditional intelligent customer service technology has hit a cognitive "ceiling". Q&A-based AI agents built on keywords or fixed workflows, due to their weak semantic understanding and insufficient contextual awareness, often fall into the dilemma of "irrelevant answers". They are unsuitable for complex and ever-changing business consultation senarios, nor can they effectively alleviate the workload of human staff.

On the other hand, the application of emerging general LLMs in vertical fields faces the severe test of the "last mile." Although their language capabilities are strong, directly applying them to serious human resources and social security policy services presents four core problems:

01 Risk of accuracy "hallucination"

General models, when unconstrained, tend to generate factually incorrect content, which is an insurmountable red line in policy consultations where errors are unacceptable.

02 Difficulty in real-time knowledge adaptation

The general knowledge repository built into models cannot synchronize in real-time with the frequently updated and differentiated policy details across regions, inherently suffering from lag.

03 Lack of process traceability

Their "black box" nature makes it difficult to trace output answers back to specific policy clauses, failing to meet the authoritative and trustworthy requirements of government services.

04 Insufficient content security and controllability

Ensuring the model's output is always legal, compliant, secure, reliable, and strictly within service boundaries is key to safeguarding public service security.

Therefore, how to overcome the cognitive bottlenecks of traditional technology while effectively solving the inherent defects of general LLMs in terms of accuracy, synchronization, controllability, and traceability has become the core challenge that must be overcome for the current intelligent upgrade of human resources and social security services.

SOLUTION AND VALUE

01 Overall Structure/Solution

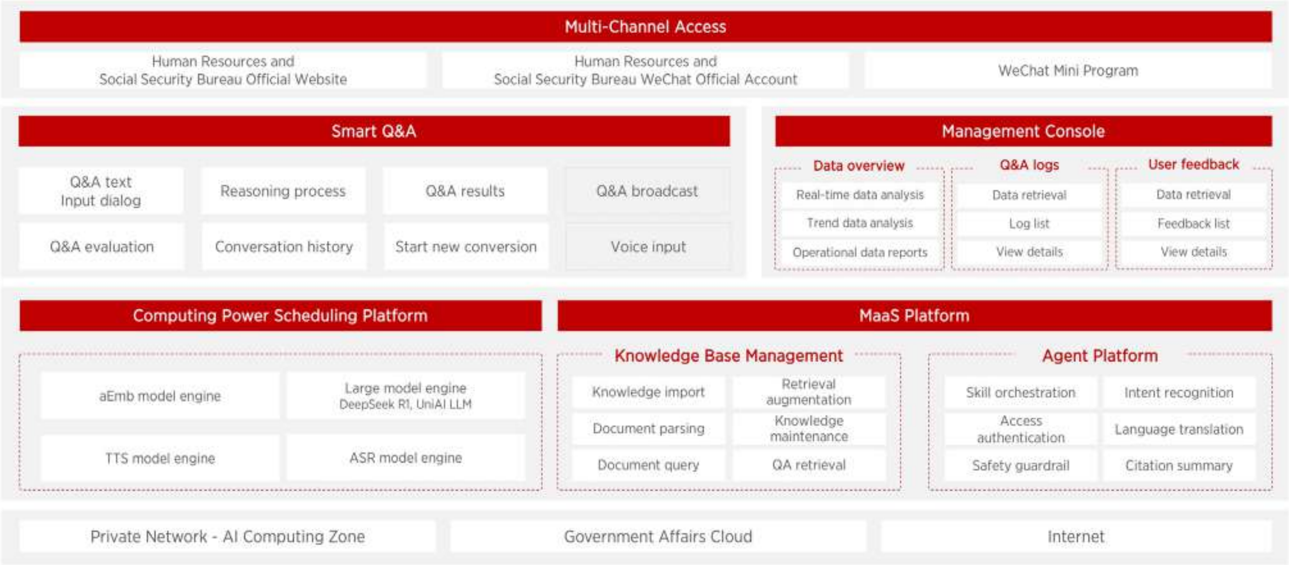

Multi-Channel Access
Human Resources and Social Security Bureau Official Website
Human Resources and Social Security Bureau WeChat Official Account
WeChat Mini Program
Smart Q&A
Q&A text Input dialog
Reasoning process
Q&A results
Q&A broadcast
Q&A evaluation
Conversation history
Start new conversion
Voice input
Management Console
Data overview
Real-time data analysis
Trend data analysis
Operational data reports
Q&A logs
Data retrieval
Log list
View details
User feedback
Data retrieval
Feedback list
View details
Computing Power Scheduling Platform
aEmb model engine
Large model engine DeepSeek R1, UniAI LLM
TTS model engine
ASR model engine
MaaS Platform
Knowledge Base Management
Knowledge import
Retrieval augmentation
Document parsing
Knowledge maintenance
Document query
QA retrieval
Agent Platform
Skill orchestration
Intent recognition
Access authentication
Language translation
Safety guardrail
Citation summary
Private Network - AI Computing Zone
Government Affairs Cloud
Internet

Figure 1: System architecture

The system supports multiple computing power sources. Based on the computing power, the scheduling platform and MaaS platform enable the deployment and management of underlying models. The intelligent Q&A system comprehensively manages knowledge content, answering strategies, and conversation history to respond to citizens' inquiries.

Intelligent interactive server

① **Multi-channel/multi-terminal adaptation:** Embedded in local human resources and social security official accounts, official websites, and the "People's Livelihood One-Card" mini program, supporting a unified Q&A entry point across multiple terminals (mobile/PC).

② **Multiple input methods:** Supports text and voice input to meet the needs of different user groups.

③ **Multi-turn dialog and context understanding:** Supports multi-turn conversations and context understanding, providing a coherent interactive experience capable of handling complex consultations requiring multiple interactions.

④ **Accurate interpretation of policy documents:** The robot's answers accurately match the original policy text, are traceable to sources, and provide Q&A combined with text and images, effectively ensuring the compliance and authority of policy interpretation, and delivering efficient and reliable intelligent policy consultation services.

⑤ **Customized training for complex scenarios:** Customized training for complex scenarios such as retirement age calculation, providing more user-friendly and accurate Q&A services.

⑥ **Closed-loop service:** Can transfer to manual agents when complex or personalized requests cannot be met.

Agent management and knowledge governance

① **Agent management:** Enables unified management of agents, API authorization, log management, data access, model training, etc.

② **Knowledge base management:** Supports batch import of knowledge bases and multiple document types.

③ **Knowledge base classification processing:** Classifies the knowledge base. Publicly available policy documents are referenced and shown to citizens in the "Jing Xiao Bao" Q&A; documents not suitable for public release, such as internal guidelines, FAQs, and policy interpretations, are for the LLMs' reference only and are not cited or displayed.

Global monitoring and alerting

① **User feedback collection:** Users can evaluate the system service and voice reasons for dissatisfaction; the system records data for iterative optimization.

② **Intelligent monitoring dashboard:** Monitors real-time system operational data, generates operational reports, analyzes conversation trends, and provides early warnings for hot issues.

③ **System full-link monitoring:** Implements full-link monitoring for the "Jing Xiao Bao" system, providing real-time alerts for abnormal key indicators to ensure system stability.

02 Application Scenarios

Scenarios	Solutions	Cases
Daily high-frequency policy queries	24/7 availability, providing second-level responses and accurate answers to standardized, high-frequency questions. Local human resources and social security official accounts effectively handle 60%-80% of standardized inquiries daily, equivalent to adding 5-8 human agents to the hotline. Complex personalized business calculations	The official social media account of a local Human Resources and Social Security Bureau effectively responds to various high-frequency inquiries daily. Over 60% of these inquiries are handled by the online assistant, replacing manual responses. Based on local labor costs, this translates to an estimated reduction of 50,000 to 100,000 yuan in monthly expenses.
Complex personalized business calculations	Specialized training based on individual parameters, providing accurate calculation services for complex businesses like retirement age, social security allowances, and unemployment insurance benefits.	Used by multiple municipal human resources and social security bureaus; accuracy rate for personal business calculations (pension/unemployment benefits) > 99.5%, replacing previously cumbersome manual calculations.
New policy release & intensive publicity	After a new policy is released, the knowledge base is updated within hours, ensuring unified, authoritative, and timely external communication.	On the first day of a new medical insurance policy release in a certain city, it smoothly handled a consultation peak over 5 times the normal volume, with an average user wait time of less than 5 seconds.
Internal empowerment& knowledge management	Serves as a "policy knowledge co-pilot" for frontline staff, providing instant and accurate policy queries and reference guidelines, improving internal work efficiency.	At a district social security center, the training cycle for new employees before independent work was shortened by 30%, and internal policy query consistency increased to over 98%.

Figure 2: Application effect monitoring

In practical application, the system has established multi-dimensional indicator monitoring to ensure stable system operation and real-time feedback on Q&A effectiveness.

03 Key Innovations

Deep domain fine-tuning

Uses SFT technology for targeted model optimization, increasing policy interpretation accuracy from 61% to 88%, ensuring strictly compliant model output.

Intelligent knowledge base scheduling

Separate storage by database + priority retrieval strategy, combined with cross-business vector retrieval, significantly improves policy recall rate and reduces latency.

Policy scenario simulator

Automatically calculates social security amounts—such as pensions—based on user input parameters, generating charts + interpretations in real-time, achieving "one-click calculation" for policies.

Cloud-edge collaboration

Uses a distributed architecture to keep sensitive data processing within the government cloud, deploying only lightweight frontends on the Internet, meeting the requirement that government data does not leave the domain.

Elastic scaling

The computing power scheduling platform supports online expansion. During surges in consultation volume, computing power servers can be rapidly scaled up to quickly respond to high concurrency demands.

- Proactive inquiries & counter-questions to complete information: Automatically triggers follow-up questions for vague queries, increasing the conversion rate for complex business consultations by 65%, ensuring answer accuracy.
- Convenient voice Q&A entry: Long-press for voice input to ask questions directly, eliminating typing for elderly users, achieving barrier-free "ask by speaking" service.

04 Business Models

Model	Charging method	Target customers	Cases
Private deployment	One-time software system licensing fee, plus an annual maintenance service fee.	Large government departments or enterprises with their own servers and high-security requirements.	Internal knowledge base for a local HRSS bureau.
All-in-one terminals	One-time purchase of hardware equipment including pre-installed software, with optional annual maintenance renewal.	Government departments at various levels seeking rapid deployment and simplified IT operations.	Q&A assistant for a municipal HRSS bureau.
System engineering & customized services	Project-based packaged pricing, or billing based on expert effort invested.	Clients with specific needs for system integration, functional development, or long-term outsourced operations.	Provincial (or municipality) level HRSS Q&A assistant.

05 Core Values

By utilizing DeepSeek + RAG technology, it realizes a LLM-based intelligent Q&A system specifically for the vertical field of human resources and social security. This provides a replicable model for the digital transformation of HRSS systems nationwide and other government service systems.

It currently covers over 1,200 active, publicly available policy documents, more than 8,000 frequently asked questions (FAQs), approximately 2,700 curated Q&A pairs, and around 6,800 core knowledge points. Data sources include internal policy databases from local municipal HRSS bureaus, FAQ libraries from various 12333 hotlines, and business handling guides provided by subordinate units of the HRSS bureaus. This knowledge encompasses areas like social insurance, employment and entrepreneurship, labor relations, and personnel examinations.

1200

active, publicly available policy documents

8000

frequently asked questions (FAQs)

2700

curated Q&A pairs

6800

core knowledge points

The Q&A application supports multi-turn conversations, follow-up questions, and clarifications across HRSS domains like social insurance, employment and entrepreneurship, labor relations, and personnel examinations, enabling natural and fluent responses to citizen inquiries.

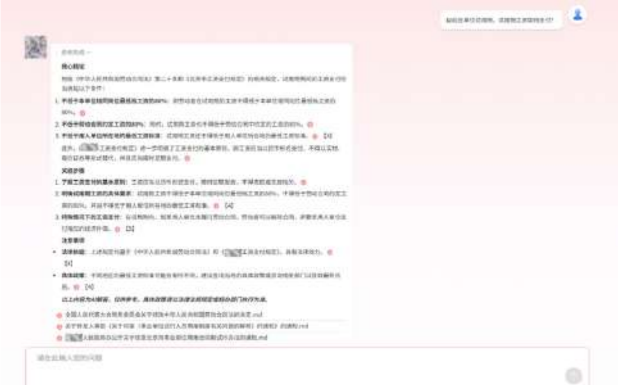
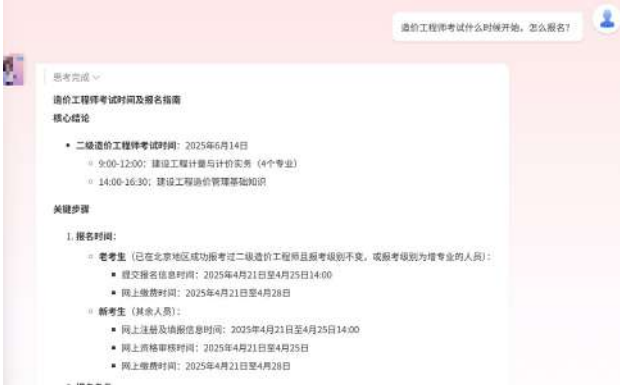


Figure 3: Employment-related Q&A example



造价工程师考试什么时候开始，怎么报名?

思考完成

造价工程师考试时间及报名指南

核心结论

- **二级造价工程师考试时间**：2025年6月14日
  - 9:00-12:00：建设工程计量与计价实务（4个专业）
  - 14:00-16:30：建设工程造价管理基础知识

关键步骤

1. **报名时间**：
   - **老考生**（已在北京地区成功报考过二级造价工程师且报考级别不变，或报考级别为增专业的人员）：
     - 提交报名信息时间：2025年4月21日至4月25日14:00
     - 网上缴费时间：2025年4月21日至4月28日
   - **新考生**（其余人员）：
     - 网上注册及填报信息时间：2025年4月21日至4月25日14:00
     - 网上资格审核时间：2025年4月21日至4月25日
     - 网上缴费时间：2025年4月21日至4月28日

Figure 4: Personnel examination-related Q&A example

REFLECTION

Key Insights

Scenario-driven, targeted problem-solving	Focusing on high-frequency, complex scenarios like pension calculation and unemployment insurance claims as the core entry points, we specifically constructed multi-turn dialog and specialized algorithm modules. This approach effectively handles nearly 90% of standardized inquiries, successfully validating the practical implementation path of "focusing on core pain points and optimizing through iterative steps," ensuring maximum value from the technology investment.
Technology integration, deeply understanding needs	Adhering to the principle that technology serves application. We adopted the RAG architecture to ensure over 95% of answers can be accurately traced back to the original policy text, guaranteeing service authority; developed specialized algorithm modules to solve the public's pain point of "calculation difficulty"; and utilized SFT to achieve accurate translation of policy language into colloquial expressions, making the service truly accessible to the public.
Knowledge middle platform, dynamic empowerment	A multi-source, heterogeneous knowledge hub for HRSS was built, breaking down information silos through a knowledge association network. An agile response mechanism enables knowledge base synchronization and guideline unification within 72 hours of policy release, ensuring the assistant's knowledge system always remains "fresh" and accurate.
Closed-loop evolution, continuous learning	An AI self-optimization closed loop of "service-monitoring-learning" was established. The system can actively learn from manual corrections and, through cluster analysis of hot issues, intelligently warn of potential knowledge gaps, driving the iterative improvement of the knowledge base, achieving continuous evolution where it "gets smarter the more it is used."

Future Plans

Multi-turn interactive Q&A training

Continue investing in specialized training for multi-turn interactions within complex business scenarios. This aims to further enhance the model's capabilities in context tracking and long-range logical reasoning, ensuring that the accuracy and logical consistency of responses reach new heights when handling multi-step, interdependent consultation tasks.

12333 hotline voice assistant interactive service

Implementation of the AI voice assistant for the 12333 hotline. By introducing advanced speech recognition and synthesis technologies, we aim to create a "never busy" AI voice navigator. This will not only efficiently divert call volume pressure from human agents but also cater to the usage habits of different groups, including the elderly, achieving comprehensive coverage and inclusive access to service channels.

Telecom LLM-enabled Digital Intelligence Products

In the information era, the telecom industry plays a crucial role as the infrastructure of communication networks. Facing massive user data, complex network structures, and diverse business needs, the telecom industry has to build a large language model that helps operators make their business models more intelligent and autonomous.

Peng Li
GUOCHUANG Software Co., Ltd.
Vice President of Telecom Operators BG

SOLUTION PARTNERS 科大国创 GUOCHUANG

Large language models, with their powerful computational capabilities and learning abilities, can extract valuable information from massive data, providing telecom operators with precise user profiling, optimized network layout, improved service quality, and more. Moreover, the telecom industry can achieve intelligent operations with lower costs and higher efficiency. Telecom operators can make their business models more intelligent and autonomous, empowering scenarios such as intelligent ticket scheduling, intelligent network O&M, and intelligent customer service.

Based on massive operator data accumulated over time, GUOCHUANG integrates key technologies such as knowledge graph enhancement and few-shot learning in the operator LLM to solve technical challenges in industry applications, such as difficult knowledge injection and poor generalization. Through domestically controlled multi-granularity open knowledge graphs and multimodal pre-trained LLM bases, it forms an independently innovative knowledge-enhanced LLM framework, empowering three business scenarios: intelligent ticket scheduling, intelligent network O&M, and intelligent customer service. Targeting smart customer service, operations scheduling, and cloud network collection and control, GUOCHUANG innovatively presented ChatAI, ChatTicket, and ChatNet, realizing human-like front-end service, unattended back-end operation, and underlying intelligence.

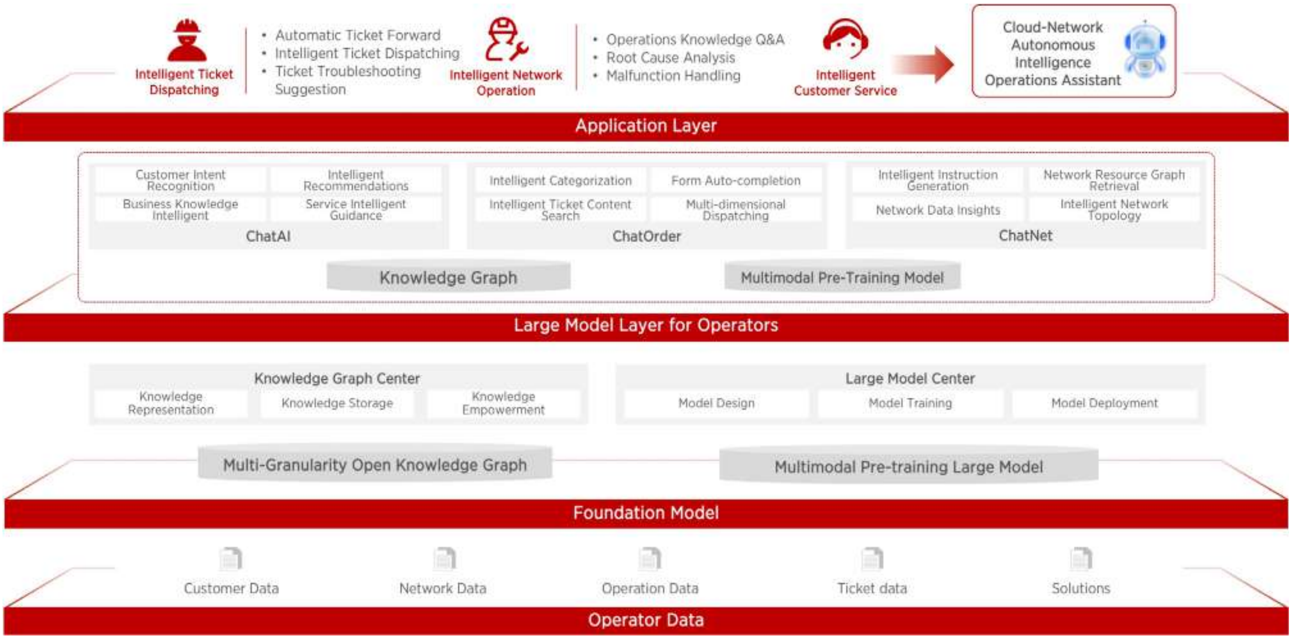

Intelligent Ticket Dispatching
Automatic Ticket Forward
Intelligent Ticket Dispatching
Ticket Troubleshooting Suggestion
Intelligent Network Operation
Operations Knowledge Q&A
Root Cause Analysis
Malfunction Handling
Intelligent Customer Service
Cloud-Network Autonomous Intelligence Operations Assistant
Application Layer
Customer Intent Recognition
Intelligent Recommendations
Business Knowledge Intelligent
Service Intelligent Guidance
ChatAI
Intelligent Categorization
Form Auto-completion
Intelligent Ticket Content Search
Multi-dimensional Dispatching
ChatOrder
Intelligent Instruction Generation
Network Resource Graph Retrieval
Network Data Insights
Intelligent Network Topology
ChatNet
Knowledge Graph
Multimodal Pre-Training Model
Large Model Layer for Operators
Knowledge Graph Center
Knowledge Representation
Knowledge Storage
Knowledge Empowerment
Large Model Center
Model Design
Model Training
Model Deployment
Multi-Granularity Open Knowledge Graph
Multimodal Pre-training Large Model
Foundation Model
Customer Data
Network Data
Operation Data
Ticket data
Solutions
Operator Data

Figure 1: Overall Architecture

CHALLENGES

Business support

Customer service, integrated coordination, and collection and control involve many scenarios with different granularity levels and severe fragmentation. It is difficult to support each fragmented scenario individually through traditional manual methods, which brings efficiency and cost pressure.

Data silo

Data is fragmented across various operator business systems, including billing systems, network management platforms, CRM systems, etc., with daily utilization of PB-level data less than 15%. Statistics from a provincial operator show that cross-department data retrieval takes more than 72 hours on average.

Inefficient network O&M

When processing millions of network element alarms, traditional rule engines are less accurate in fault localization, with an average fault recovery time of about 4 hours, resulting in low network O&M efficiency.

Service capability bottlenecks

Manual customer service handles approximately 15 tickets/person/day, with customer waiting times exceeding 30 minutes during peak periods. User satisfaction consistently hovers around 78%, which is relatively low.

SOLUTION AND VALUE

01 Overall Structure/Solution

Overall architecture

Based on three intelligent platforms, the overall solution creates three assistants to form an integrated knowledge computing system. The Intelligent Trusted Development Platform is used for trusted development capabilities, including production process orchestration and packaged business capabilities (PBC) to quickly support the three businesses above. The Intelligent Agent Platform integrates and manages standardized AI capabilities to assist the businesses above in building intelligent agents and rapidly implementing AI capabilities in business scenarios. The Intelligent Data Element Service Platform is designed for the retrieval, integration and management of data elements, providing stable "data" for the three major businesses above.

Through such steps as scenario rule organization, data processing and model development, intelligent agent construction, system integration, and operational iteration, it efficiently supports operators in business processes for customer service, scheduling, and network operation.

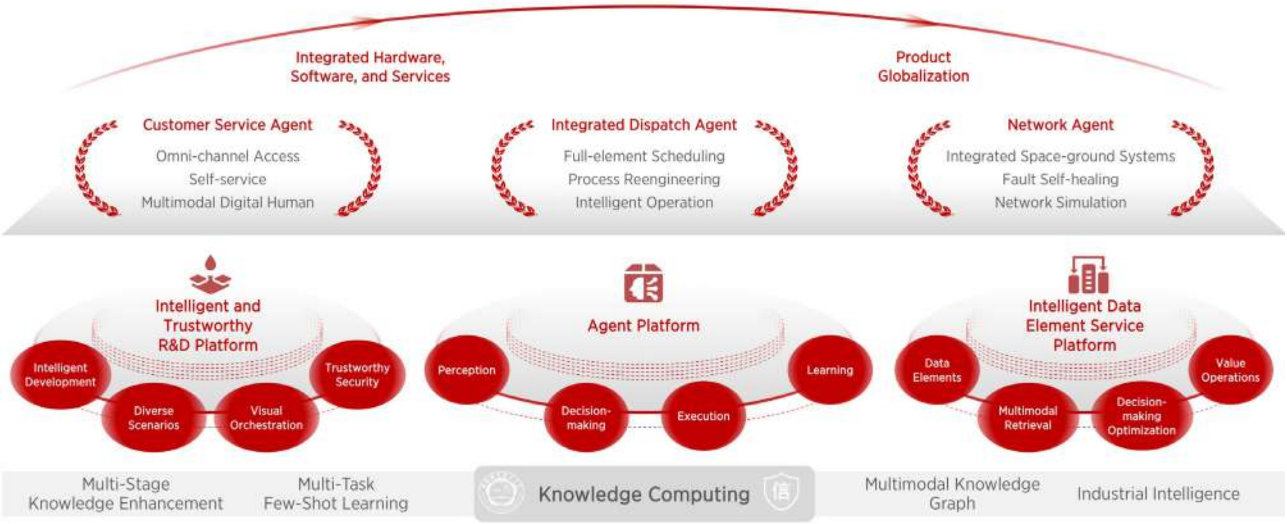

Integrated Hardware, Software, and Services
Product Globalization
Customer Service Agent
Omni-channel Access
Self-service
Multimodal Digital Human
Integrated Dispatch Agent
Full-element Scheduling
Process Reengineering
Intelligent Operation
Network Agent
Integrated Space-ground Systems
Fault Self-healing
Network Simulation
Intelligent and Trustworthy R&D Platform
Intelligent Development
Diverse Scenarios
Visual Orchestration
Trustworthy Security
Agent Platform
Perception
Decision-making
Execution
Learning
Intelligent Data Element Service Platform
Data Elements
Multimodal Retrieval
Decision-making Optimization
Value Operations
Multi-Stage Knowledge Enhancement
Multi-Task Few-Shot Learning
Knowledge Computing
Multimodal Knowledge Graph
Industrial Intelligence

Figure 2: Overall Solution

Technical solution

The technical solution includes three parts: the operator LLM, three intelligent platforms, and three business agents.

Operator LLM	The development of operator LLM mainly includes: business data collection, training corpus construction, fine-tuning corpus construction, LLM pre-training, LLM instruction fine-tuning, LLM evaluation and optimization, LLM inference acceleration, open LLM services, and AI-assisted SLM capacity building for three major scenarios — telecom customer service, operations scheduling, cloud network collection and control.
Three intelligent platforms	The development of three intelligent platforms mainly includes: the development and deployment of the Intelligent Trusted Development Platform, Intelligent Agent Platform, and Intelligent Data Element Service Platform.
Three business agents	The development of three business agents mainly includes: the development and deployment of intelligent customer service agents, intelligent ticket agents, and intelligent network agents. Correspondingly, it means the development and application of three intelligent products: ChatAI, ChatTicket, and ChatNet.

02 Application Scenarios

ChatAI

The intelligent customer service system enables human-like interaction through process orchestration, knowledge modeling, and data integration. It supports natural language dialog, real-time decision assistance, and automated operations. The service process is divided into two phases: self-service support and human agent assistance. The LLM technology enables collaboration among voice navigation, virtual digital humans, and text robots. When human agents are required, a multimodal system analyzes user intent, identifies issues, and recommends solutions, significantly improving service efficiency and quality.

ChatTicket

Based on multi-modal interaction and real-time data retrieval, the intelligent installation and maintenance assistant integrates on-site guidance, equipment parameter database, fault maintenance, and marketing knowledge graph, providing technical support and scheduling services for engineers and backend teams. It also resolves complex on-site issues through natural language interaction. The intelligent ticket agent architecture enables automated ticket management throughout the entire lifecycle. By building a digital knowledge base and leveraging NLP for automatic classification and priority determination, it supports ticket creation, review, dispatching, handling, and completion. The ticket transfer process is monitored in real time for abnormal warnings.

ChatNet

ChatNet, with its twin topology engine as the core, can rapidly fix faults in cloud network. Through a symbiotic multicloud architecture and ultra-fast simulation, it can effectively tackle challenges in resource management. The system enables fully intelligent fault management: detecting faults within 1 minute, identifying root causes within 5 minutes, and automatically handling issues within 10 minutes. It guarantees a better and more efficient network O&M experience.

03 Key Innovations

Technical innovation

A dedicated knowledge base for telecom operators is built by combining RAG and multimodal perception technologies, empowering intelligent parsing of text, voice, ticket screenshots, and interactive information in diverse formats. In this way, it can precisely understand user intent and generate scenario-based solutions based on LLM reasoning capabilities.

A cascade serving system is integrated with MOE (Mixture of Experts) models to build an intelligent decision-making engine covering the entire ticket process. The three mechanisms of ticket intent recognition, dynamic resource matching, and execution path optimization enable AI-driven end-to-end automatic ticket transfer.

Business innovation

For online customer service

the traditional passive response model has been revolutionized. Customer requests are generated through interactive dialog and automatic ticket recognition is triggered through backend assistants. The customer service scenario is upgraded from "problem-solving" to "early intervention + process closure".

For ticket scheduling

a dialog-based ticket scheduling interface is created. Frontline personnel can trigger automatic ticket creation, resource preemption, progress synchronization, and assisted execution through natural language commands, driving the O&M mode from "multi-level collaboration" to "simplified individual operations".

For network O&M

the large model automatically captures network anomalies and associates historical fault databases with real-time performance data for root cause analysis. At the same time, the system generates integrated solutions that include processing steps and a list of spare parts. As a result, it makes experience-dependent network O&M intelligent.

Project highlights

Deep AI integration into the entire business process, comprehensively empowering the operator's business system

Innovation in AI technology drives a qualitative leap in operator services. Through deep integration with core business scenarios such as customer service responses, ticket transfer, network monitoring and troubleshooting, the system realizes the intelligent understanding of user needs, accurate identification of root causes, and autonomous decision-making for handling strategies. It fundamentally transforms the operator's service system from “AI-assisted” to “AI+driven”.

Intelligent multi-modal interaction, revolutionizing customer service and O&M experiences

Multi-modal interaction involving text, voice, images, and videos reshapes customer service interfaces and internal O&M models. It provides natural, convenient, and secure service interactions for end users; offers efficient and intuitive assistance tools for customer service agents and O&M engineers, significantly improving service response time, customer satisfaction, and problem-solving efficiency.

End-to-end process intelligence, driving leaps in service efficiency and quality

Intelligent LLM-driven guidance of critical business processes enables automated handling of clearly defined, highly repetitive tasks. Complex operational procedures are simplified and standardized, significantly improving the efficiency and first-time resolution rate of service activation, fault repair, and complaint resolution.

Early complaint and ticket processing, building an efficient closed-loop handling mechanism

AI technology is utilized for intelligent preliminary processing of massive customer complaints and generated tickets, achieving intelligent problem diagnosis, automatic root cause analysis, precise impact scope identification, and automated handling in certain scenarios. It significantly improves ticket transfer efficiency, first-time resolution rate, and customer satisfaction, while substantially reducing repetitive judgments for frontline staff.

A global intelligent knowledge hub to accumulate experience and empower full-scenario decision-making

Structured and unstructured trusted knowledge sources across all business domains (product pricing, service policies, network standards, troubleshooting manuals, historical cases, ticket records, etc.) are integrated to build a unified, dynamically updated expert-level knowledge base. This ensures that the large model outputs highly accurate, consistent, and actionable information in customer service Q&A, ticket assistance, O&M decision-making, and other scenarios, truly transforming enterprise knowledge and experience into core productivity.

04 Business Models

① Increased work efficiency

Previously, manual customer service took 5-10 minutes to handle complex inquiries. With LLM, simple questions can be answered within seconds, and the handling time for complex issues has been reduced to 2-3 minutes. It can decrease human input by 30%, significantly reducing labor and operational costs.

For common network troubleshooting, large models can reduce the average handling time from hours to minutes, improving troubleshooting efficiency by over 65%, and decreasing human input by 20%.

② Optimized workflow

The optimized ChatAI process is reflected at three levels. On the customer side, the customer service assistant can quickly respond to various requests such as inquiries, complaints, and technical issues, providing intelligent processing solutions; on the agent side, it provides business guidance, assistance, and proxy services, effectively improving workflow efficiency; on the management side, it supports service analysis, service early warning and other decision-making processes, effectively improving service efficiency.

ChatTicket helps with business process execution, filling out ticket pages through intelligent recommendations and automatic filling, or creating tickets through intelligent email recognition, etc. During the process, interactive dialogs are conducted for the system to precisely retrieve tickets, recommend handling suggestions, and provide guidance for ticket handling. Alternatively, digital employees are utilized to initiate automated handling and generate relevant summaries and reports. In business process operations, compliance checks and deviation detection are conducted through data mining and analysis to identify bottlenecks and areas of improvement, thereby assisting in business process optimization.

ChatNet revolutionizes the traditional O&M model characterized by multi-department collaboration, lengthy processes, and low efficiency and makes the process autonomous for self-repair. The intelligent agent can fully leverage its perception and analysis advantages to ensure faults are detected within 1 minute. Through comprehensive intelligent reasoning and positioning, the root causes can be quickly identified within 5 minutes. With its intelligent decision-making and repair capabilities, the agent can quickly generate and implement fault resolution plans within 10 minutes, achieving agile self-repair of network failures.

REFLECTION

Key Insights

The three LLM-enabled digital intelligence products have been implemented in projects of China Mobile, China Telecom, and China Unicom, as well as their provincial companies, greatly enhancing the intelligence level of the telecom industry. Thanks to them, operators can realize large-scale, efficient, digital, and intelligent applications of business knowledge and data, garnering significant economic value and social benefits.

With the existing knowledge-enhanced LLM architecture and capabilities, GUOCHUANG's operator LLM and three intelligent products are being used in more industries like transportation, energy, and smart cities. They are tightly coupled with business systems through the three intelligent platforms, realizing rapid cross-provincial deployment. The fast, highly-available model can significantly shorten development cycles, reduce implementation costs, and enhance industry intelligence levels through data and knowledge. As a result, the project has been successfully implemented in multiple provinces, providing a mature example for the intelligent transformation of the telecom industry.

Future Plans

Business level

Expanding business application scenarios:

- Intelligent customer service upgrade: Integrate omni-channel services throughout the customer lifecycle and provide personalized services.
- Refined operations scheduling: Integrate systems like ERP to dynamically adjust ticket allocation strategies based on multidimensional factors.
- Advanced autonomous network O&M: Autonomous network planning, deployment, and dynamic resource optimization.

Technical level

Deepening technological innovation and optimization:

- Enhance knowledge graph fusion: Improve multi-granularity knowledge graphs with emerging business knowledge; optimize fusion mechanisms to enable fast and precise knowledge retrieval.
- Improve few-shot learning: Develop advanced meta-learning algorithms to enhance model learning and adaptation capabilities with limited samples.
- Optimize model architecture: Introduce new neural network structures to improve long sequence processing; apply technologies such as model distillation to achieve lightweight edge deployment.

Industry level

Achieving ecological integration and expansion:

- Cross-industry application: Create LLMs and applications specific to industries such as energy, transportation, etc.
- Deepened industry-academia-research cooperation: Conduct technical research with universities and research institutions, and participate in industry standards development.
- Upstream and downstream collaboration: Cooperate with data, computing power, and application enterprises to build a win-win industrial ecosystem.

Large Model Project for Medical Imaging Quality Control

> **The volume of medical imaging data is surging, but the lack of quality control standards and variations in manual operations make it difficult to accumulate high-quality datasets. This has become a critical bottleneck in medical AI development and healthcare cost control. Targeting high-quality medical datasets, we focus on establishing a unified and intelligent system for medical imaging and report quality control. It ensures data consistency and standardization, eliminating low cross-institutional mutual recognition, reducing redundant scans, and lowering healthcare costs. We strengthen the data foundation for healthcare insurance reform through standardized governance and accelerate the construction of imaging mutual recognition systems. This supports the "Healthy China" initiative and digital transformation of healthcare, and promotes sustainable development of public finances.**
>
> **Chen Haifeng**
> **China Unicom Data Intelligence Co., Ltd.**
> **Chairman, General Manager**

SOLUTION PARTNERS

中国联通 China unicom | UniAI 联通元景 | 海纳医康 | 云知声 Unisound

Amid the explosive growth of medical imaging data, we take cross-institutional mutual recognition of imaging results as the entry point, actively exploring intelligent solutions to address the efficiency bottlenecks and standardization challenges of traditional quality control models.

The accuracy of medical imaging and reports has a direct bearing on clinical diagnosis. Traditional quality control, which relies on manual sampling, is grappling with inconsistent standards, high rates of missed inspections, and the sheer volume of data. Primary healthcare facilities, limited by insufficient resources, are even more reliant on high-quality radiology reports. In response, the Beijing Medical Imaging Quality Control Center, in collaboration with a technical team, has constructed an intelligent medical imaging quality control platform that integrates "data-model-application." This platform leverages cloud imaging technology, data resources from multiple institutions, and multimodal large models for image quality assessment and in-depth semantic analysis of report texts. It significantly improves the coverage and real-time capability of quality control while filling the gaps inherent in manual sampling.

This platform has achieved four breakthroughs. First, it standardizes medical imaging and radiology report output through AI quality control models, reducing the risk of misdiagnosis and missed diagnosis. Its core advantage lies in the deep integration of business scenarios with technical capabilities to form a replicable "intelligent quality control" paradigm. Second, based on big data analysis capabilities, it supports regional disease spectrum research and population health management. Third, it provides a technical foundation for telemedicine and tiered diagnosis and treatment, helping quality resources flow to primary medical institutions. Fourth, the "business-driven + technology-led" model provides practical evidence for formulating national medical imaging quality control standards. In the future, it can be replicated in areas such as cross-regional mutual recognition and scientific research data governance, driving up healthcare quality.

CHALLENGES

Medical imaging quality control is key to improving diagnosis and treatment efficiency, reducing healthcare costs, and ensuring patient safety. However, it still faces challenges in practical application, and breakthroughs through AI technology, especially large medical models, are urgently needed. Automated and standardized transformation is achieved through multimodal large models to address major pain points such as data volume, quality variation, mutual recognition barriers, and complex semantics, aiding the implementation of the national "data element standardization" strategy.

Large data volume with varying quality, resulting in inefficient manual quality control

Medical imaging data is experiencing explosive growth, but data quality varies significantly due to factors such as equipment models, operational standards, and diagnostic expertise. Traditional quality control relies on manual sampling, but a single doctor can only handle quality control tasks for fewer than 100 cases per day. The massive data means high labor costs and extremely low coverage. Furthermore, manual procedures are susceptible to subjective factors, and inconsistent standards lead to high rates of missed inspections. It is thus difficult for hospitals to build high-quality datasets.

Low rate of mutual recognition among institutions leads to repeated examinations, exacerbating resource waste
The current mutual recognition rate for medical imaging among institutions is only 25%, mainly due to equipment differences, inconsistent operational standards, and insufficient radiology report standardization. Patients have to take repeated examinations, which not only increases their financial burden but also prolongs the diagnosis and treatment cycle.

Quality control of radiology reports poses significant challenges, particularly due to pronounced semantic complexity
Radiology reports involve a large number of technical terms, logical correlations, and multimodal information. Manual review tends to overlook typographical errors, slips of the pen, or underlying logical errors. Daily report volume per person can exceed 20,000 characters. Traditional rule engines struggle to cover complex semantic scenarios, leading to risks of incorrect information transfer and affecting the accuracy of clinical decisions.

Significant limitations of traditional AI technology
Early machine learning or deep learning models rely on small sample training, leading to poor generalization and adaptation to diverse equipment models (such as 16-slice to 256-slice CT scanners) and scanning sites (over 100 types). While the era of generative AI brings new possibilities, it is necessary to address issues such as data privacy, model explainability, and seamless integration with clinical workflows.

SOLUTION AND VALUE

01 Overall Structure/Solution

Focusing on high-quality dataset construction and healthcare insurance cost control, the project aims to build a unified, intelligent medical imaging and report quality control platform. It helps promote the establishment of a standardized, intelligent medical data review system, improve the mutual recognition of imaging data, reduce medical costs, and achieve coordinated development of data governance, semantic quality control, domestic technology substitution, and standard co-construction. Furthermore, it aids the industrialization of medical AI and the implementation of national health strategies.

To achieve these goals, the project builds a technical architecture covering three levels—data, models, and applications—based on a domestic software and hardware platform.

01 The data level

The project integrates high-quality datasets of medical imaging quality control, like report data, 2D images, 3D images, and video images. This ensures consistency, standardization, and interoperability of data quality, providing a solid foundation for large model training and application.

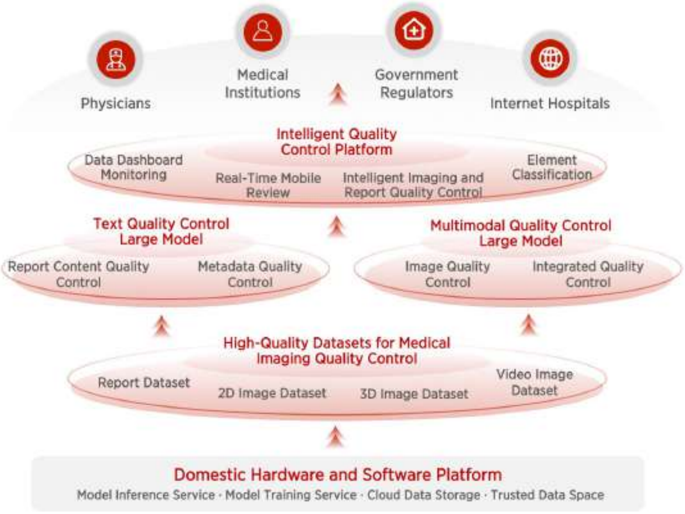
02 The model level

Large models for both text quality control and multimodal quality control have been developed. The text quality control large model focuses on automated standardization review and quality control of reports. It leverages advanced generative large language models to automatically identify and correct issues in reports. The multimodal quality control large model enables standardized and intelligent quality control of imaging data. It utilizes image preprocessing and AI analysis technologies to enhance automated monitoring capabilities of imaging data. Heterogeneous federated learning enables cross-domain model training.

03 the application level

The project, through a unified database, serves hospitals, medical institutions, government management departments, and Internet end users. It can meet multi-level, multi-scenario requirements for medical data quality control and review, and promote full-chain quality management of medical data.

Overall, this project constructs a trustworthy, standardized, and intelligent quality control system for medical imaging and radiology reports through an integrated data-model-application design, promoting the healthy and orderly development of medical AI.


Physicians
Medical Institutions
Government Regulators
Internet Hospitals
Intelligent Quality Control Platform
Data Dashboard Monitoring
Real-Time Mobile Review
Intelligent Imaging and Report Quality Control
Element Classification
Text Quality Control Large Model
Report Content Quality Control
Metadata Quality Control
Multimodal Quality Control Large Model
Image Quality Control
Integrated Quality Control
High-Quality Datasets for Medical Imaging Quality Control
Report Dataset
2D Image Dataset
3D Image Dataset
Video Image Dataset
Domestic Hardware and Software Platform
Model Inference Service · Model Training Service · Cloud Data Storage · Trusted Data Space

Establish a demonstration platform in Beijing	Create an intelligent quality control benchmark platform that can be replicated and promoted, empowering high-quality development of medical imaging and reporting.
Multimodal Quality Control Large Model	Images and reports are integrated to intelligently achieve automated quality control and multimodal consistency verification in the medical industry.
High-Quality Datasets for Medical Imaging Quality Control	Standardized quality control data collection and annotation to support multimodal model training and facilitate industry standardization.
Computing Power Infrastructure	Provide computing power, models, and high-quality data elements for model development based on China Unicom's computing power platform and data platform

Figure 1: Solution for medical imaging and report quality control based on large models

02 Application Scenarios

Based on the large model for medical imaging quality control, it focuses on three key technologies: data governance, report quality control, and cross-modal fusion, to build standardized high-quality multimodal medical datasets, thus promoting the development of medical AI and controlling healthcare insurance expenditures.

Multimodal data governance

A system for data monitoring, data annotation, and processing standards is constructed to support standardized governance of medical imaging data and large model training. It ensures controllable processes, intelligent annotation, and resource coordination, providing a high-quality data foundation for quality control and intelligent applications.

Intelligent report quality control

Large models and knowledge graphs, combined with semantic understanding, achieve automatic quality control of structured and unstructured reports, multi-dimensional consistency verification, and multi-report correlation analysis. It improves radiology report quality and intelligent review capabilities.

Intelligent cross-modal fusion

Multimodal data, including images, text, and physiological signals, are integrated for unified semantic understanding and representation to achieve image-text consistency verification and intelligent quality control. It supports high-quality management throughout the entire process from data to diagnosis.

03 Key Innovations

The imaging and report quality control large model, based on advanced AI technology and multimodal fusion capabilities, improves medical data quality from key aspects such as standardization, mutual recognition, intelligent review, and automatic error correction, aiding precision diagnosis and treatment and optimal resource allocation. In this process, the platform presents the following key innovations:

Constructs a unified multimodal quality control system that achieves "image-text-structured data" integration

(1) Pioneers a multimodal quality control large model that integrates imaging + reports + structured information for unified semantic expression and understanding to achieve consistency verification and integrated quality control of image and text data.

(2) Supports intelligent quality control throughout the entire process from imaging and image review to report generation and diagnosis. A complete data chain promotes the implementation of AI-assisted diagnosis and treatment systems.

Establishes a standardized mechanism for massive data governance to support high-quality large model training and evaluation

(1) Constructs a comprehensive multimodal data governance system covering acquisition, annotation, cleansing, and standard formulation.

(2) Supports task distribution, collaborative annotation, data audit, pre-inspection processes, etc., to ensure the quality and consistency of training data as it is equipped with an imaging cloud platform.

(3) Develops new heterogeneous federated learning technologies to enable joint training and evaluation for cross-domain data.

Builds a municipal-level demonstration platform in Beijing, forming a new paradigm for regional intelligent quality control

(1) Breaks down barriers among healthcare institutions, regulators, and enterprises to construct a collaborative quality control platform covering doctors, hospitals, government departments, and the industry chain.

(2) Shifts from the original manual sampling to automated, normalized, full-scale quality control processes driven by large models.

(3) Implements integrated supervision, service, evaluation, and feedback to form an efficient closed-loop governance mechanism.

04 Business Models

The product model of two quality control large models and one quality control platform provides intelligent decision support for medical institutions and government regulators, as shown in the figure:

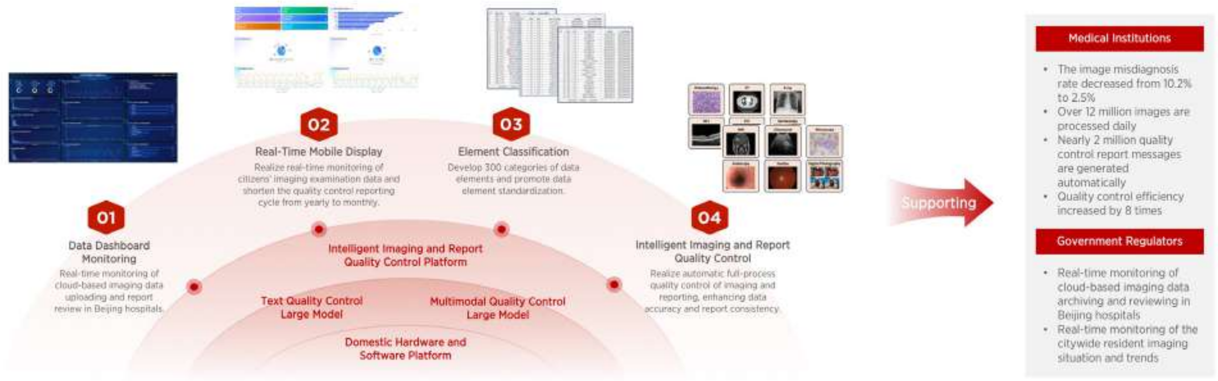

01
Data Dashboard Monitoring
Real-time monitoring of cloud-based imaging data uploading and report review in Beijing hospitals.
02
Real-Time Mobile Display
Realize real-time monitoring of citizens' imaging examination data and shorten the quality control reporting cycle from yearly to monthly.
03
Element Classification
Develop 300 categories of data elements and promote data element standardization.
04
Intelligent Imaging and Report Quality Control
Realize automatic full-process quality control of imaging and reporting, enhancing data accuracy and report consistency
Intelligent Imaging and Report Quality Control Platform
Text Quality Control Large Model
Multimodal Quality Control Large Model
Domestic Hardware and Software Platform
Supporting
Medical Institutions
The image misdiagnosis rate decreased from 10.2% to 2.5%
Over 12 million images are processed daily
Nearly 2 million quality control report messages are generated automatically
Quality control efficiency increased by 8 times
Government Regulators
Real-time monitoring of cloud-based imaging data archiving and reviewing in Beijing hospitals
Real-time monitoring of the citywide resident imaging situation and trends

Figure 2: Intelligent imaging and report quality control platform

Target customers: Targeting regional medical imaging quality control centers, while reaching secondary- and higher-level hospitals as potential customers.

Profit model: Building a closed loop of "model service + hospital access + regulatory feedback + industrial support" to drive collaborative research and development and standard construction among AI imaging enterprises. The algorithm models, platform software, hardware equipment, and networks are sold as packaged solutions, while customized services are separately charged.

05 Core Values (including quantitative results and qualitative impact)

This scenario comprehensively empowers the AI medical industry chain to promote the coordinated development of model services, hospital access, enterprise research and development, and joint standard construction. It has good prospects for promotion in more industries. The core values are as follows:

Efficient data aggregation

By building a unified data upload system, it solves challenges such as multiple institutions, heterogeneous data, and delayed uploads, ensuring real-time data control and complete transmission. Data from 230 medical institutions has been aggregated, forming over ten million cases of standardized medical imaging reports and data, with a cumulative data volume of over one petabyte. A cross-domain computing framework for datasets has also been proposed.

Shorter quality control cycles

Thanks to multimodal large models and semantic understanding technology, quality control reporting cycles have been shortened from "yearly" to "monthly". It can realize automatic quality control of millions of entries every day, with improved efficiency.

Reduced misdiagnosis rates

The feature of image-report consistency analysis helps reduce the relative error detection rate by 30%. The mutual recognition mechanism has been significantly improved, providing strong support for medical insurance expenditure control, with estimated annual savings of nearly ten million yuan.

REFLECTION

Key Insights

The success of this project demonstrates that building high-quality medical imaging datasets requires a two-pronged approach of "standardized governance + intelligent quality control." By integrating imaging and report data through multimodal large models, automated quality control can be achieved throughout the entire process from acquisition to diagnosis. The Beijing Medical Imaging Cloud Platform has validated the feasibility of the pathway: data governance foundation - AI model empowerment - industry collaboration.

Future Plans

In the next phase, we will focus on building high-quality healthcare datasets to promote coordinated development in platform capabilities, data security, and standard systems. First, we will promote compliant and efficient utilization of medical imaging data with data security as the cornerstone. Technologies for security protection and trusted sharing of medical imaging data are leveraged for safe and orderly data circulation and utilization, ensuring that "sensitive data remains within domains while data is usable but not visible." In this way, we can make the best of medical data. Second, AI-assisted medical imaging quality control standards are developed for industry leadership. While accumulating group standards for AI-assisted medical imaging quality control, we need to play an active role in the development of local, industry, and national standards for iterative upgrades and nationwide promotion. This is how we can support high-quality development of medical imaging with standardized, scientific, and systematic standards.

Building Smart Hospitals Based on Medical Large Models

> The AI era presents a mix of opportunities and challenges in building smart hospitals. The First Affiliated Hospital of Fujian Medical University has worked with China Mobile Chengdu Institute of Research and Development ("Chengdu Institute") to create an intelligent platform covering pre-admission, in-hospital, and post-discharge services based on medical large models. It can facilitate intelligent full-cycle health management for multiple diseases with improved efficiency and personalized services. Fuyi Xiaolingtong, an innovative AI-assisted escort service jointly developed by both parties, provides patients with 7×24-hour professional support and improves their medical experience. In the future, we will explore more application scenarios to promote the development of smart healthcare ecosystems.
>
> Zheng Wei
> Director of Telemedicine and Internet Medicine Center
> First Affiliated Hospital of Fujian Medical University

SOLUTION PARTNERS
中国移动 China Mobile | 福建医科大学附属第一医院 THE FIRST AFFILIATED HOSPITAL OF FUJIAN MEDICAL UNIVERSITY

To implement the decisions and plans by the CPC Central Committee and the State Council regarding the AI+ initiative, the National Health Commission, the National Administration of Traditional Chinese Medicine, and the National Administration of Disease Prevention and Control have developed the "Reference Guidelines for Artificial Intelligence Application Scenarios in the Healthcare Industry", which clearly defines scenarios such as intelligent medical escort, intelligent medical consultation, intelligent follow-up, and AI-assisted clinical decision-making for specific diseases. Based on insights into national policies and future hospital development, the First Affiliated Hospital of Fujian Medical University, in collaboration with Fujian Mobile and Chengdu Institute, has incorporated medical large models in key processes, including guided triage, pre-consultation, post-operative follow-up, report interpretation, and digital escort assistance. They have initially established an intelligent medical service framework covering "pre-admission, in-hospital, and post-discharge" stages. The project includes full-cycle management for typical diseases such as gastric cancer, intestinal cancer, and neurological glioma, as well as a hospital-specific intelligent health assistant. In this way, they can build a new generation of smart hospitals based on AI large model technology to meet the hospital's actual needs.

CHALLENGES

Cumbersome and inefficient processes

The current medical treatment process has been fragmented by the service model and organizational structure. Patients have to make multiple trips to different windows or repeatedly log in to self-service terminals/apps for registration, examination, and medication collection, which is particularly difficult for elderly patients. An "experiential" supervision by a municipal commission for discipline inspection and supervision found that 30% of problems concentrated in process redundancies (e.g., necessity to queue up multiple times and lack of guidance).

Low utilization of auxiliary service systems

Hospitals have built diverse auxiliary service systems for patients on the Internet platform, including hospital navigation, online follow-up/consultation, health education, etc. However, due to limitations in interaction forms, these systems are often not easy to use and their utilization rates are low. The statistics show that active users of online follow-up/consultation at a certain Grade-A tertiary hospital account for only 15%-20% of outpatient visits.

High pressure on hospital services

Large Grade-A tertiary hospitals deal with a large inflow of patients from across the province and even the entire country. Patients face such a situation: long queuing time for registration, long waiting time for consultation, long queuing time for examination/payment/medication collection, and short treatment time, meaning insufficient face-to-face time with doctors.

Disconnected outpatient management

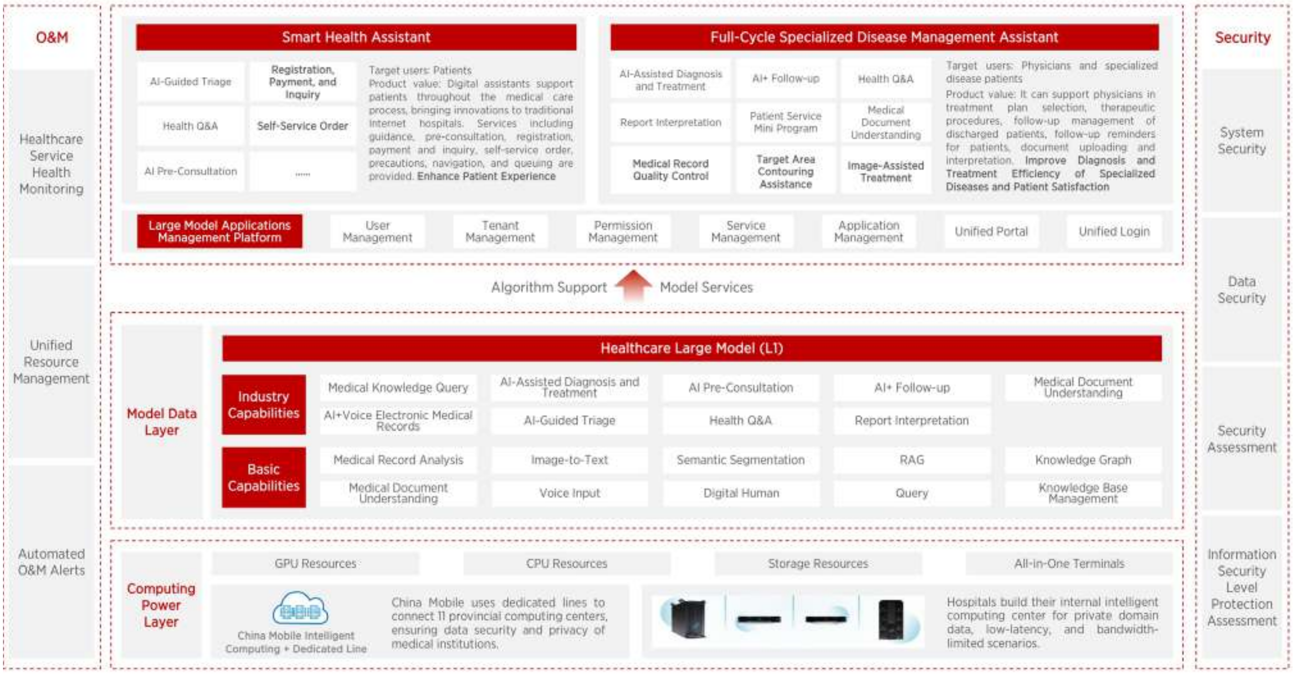
Some diseases, especially chronic conditions, require long-term follow-up visits, but current in-hospital systems are less effective in managing patients outside the hospital, with lost-to-follow-up rates exceeding 30%. Young patients are poorly compliant while elderly groups refrain from using apps due to operational difficulties, leading to interrupted health data.

SOLUTION AND VALUE

01 Overall Structure/Solution

In response to pain points in the aforementioned medical service scenarios, including patient service and full-cycle disease management, China Mobile has joined forces with the First Affiliated Hospital of Fujian Medical University to develop a large model of intelligent health assistants and full-process disease management. The model features digital human voice interaction, intelligent service agent, AI-guided triage, AI pre-consultation, information inquiry, registration and payment service, smart follow-up, and intelligent Q&A for specific diseases. With the model, the hospital has improved its ability and efficiency in patient treatment and management while meeting standards for smart services.

Built on China Mobile's "Jiutian Large Model" and DeepSeek, the Jiutian medical large model has been trained on 560G original medical data accumulated by China Mobile in the medical industry over the past years. It contains a high-quality medical industry dataset of over 40 million documents, including medical textbooks, industry papers, clinical guidelines, drug instructions, electronic medical records, etc. By selecting diverse, balanced theoretical and clinical data for training and optimization, it has forged several AI core capabilities that exceed the industry's mainstream large models, as shown in the figure.


O&M
Healthcare Service Health Monitoring
Unified Resource Management
Automated O&M Alerts
Smart Health Assistant
AI-Guided Triage
Registration, Payment, and Inquiry
Health Q&A
Self-Service Order
AI Pre-Consultation
......
Full-Cycle Specialized Disease Management Assistant
AI-Assisted Diagnosis and Treatment
AI+ Follow-up
Health Q&A
Report Interpretation
Patient Service Mini Program
Medical Document Understanding
Medical Record Quality Control
Target Area Contouring Assistance
Image-Assisted Treatment
Large Model Applications Management Platform
User Management
Tenant Management
Permission Management
Service Management
Application Management
Unified Portal
Unified Login
Algorithm Support
Model Services
Model Data Layer
Healthcare Large Model (L1)
Industry Capabilities
Medical Knowledge Query
AI-Assisted Diagnosis and Treatment
AI Pre-Consultation
AI+ Follow-up
Medical Document Understanding
AI+Voice Electronic Medical Records
AI-Guided Triage
Health Q&A
Report Interpretation
Basic Capabilities
Medical Record Analysis
Image-to-Text
Semantic Segmentation
RAG
Knowledge Graph
Medical Document Understanding
Voice Input
Digital Human
Query
Knowledge Base Management
Computing Power Layer
GPU Resources
CPU Resources
Storage Resources
All-in-One Terminals
China Mobile Intelligent Computing + Dedicated Line
Security
System Security
Data Security
Security Assessment
Information Security Level Protection Assessment

Target users: Patients
Product value: Digital assistants support patients throughout the medical care process, bringing innovations to traditional Internet hospitals. Services including guidance, pre-consultation, registration, payment and inquiry, self-service order, precautions, navigation, and queuing are provided. **Enhance Patient Experience**

Target users: Physicians and specialized disease patients
Product value: It can support physicians in treatment plan selection, therapeutic procedures, follow-up management of discharged patients, follow-up reminders for patients, document uploading and interpretation. **Improve Diagnosis and Treatment Efficiency of Specialized Diseases and Patient Satisfaction**

China Mobile uses dedicated lines to connect 11 provincial computing centers, ensuring data security and privacy of medical institutions.

Hospitals build their internal intelligent computing center for private domain data, low-latency, and bandwidth-limited scenarios.

Figure 1: Empowerment of Jiutian medical large model

02 Application Scenarios

The Jiutian medical large model can be applied in multiple processes, including medical escort, intelligent triage, pre-consultation, self-service, health education, knowledge Q&A, follow-up visits, etc., making the diagnosis and treatment workflow more intelligent.

03 Key Innovations

Service model innovation

Key large model technologies such as speech recognition, speech synthesis, and intelligent agents are leveraged to integrate various hospital systems centered on patient services. As a result, hospitals can better serve elderly patients through dialog-based interactive experiences, provide patients with excellent services, and put underutilized auxiliary services into use.

Enhanced post-discharge management

With knowledge bases, text-based large language models, and multimodal large models, it enables intelligent adjustment, automatic notifications, and intelligent result analysis for follow-up visits. It also deepens patient engagement and follow-up result statistics and extraction; strengthen post-discharge management to reduce patient loss to follow-up.

Intelligent diagnosis process

Targeting specialized disease consultation scenarios, the system integrates knowledge base, text-based large language models, and multimodal large models to cover key stages, including pre-surgery inquiry, post-surgery follow-up analysis, in-hospital medical record and examination interpretation, and medical report recognition. Through intelligent management of each step in the diagnosis and treatment process with large models, it significantly reduces doctors' manual input and shortens patient waiting time, making doctor-patient interaction more efficient.

Intelligent task guidance

The model is deeply integrated with hospital information systems to obtain patient visit status in real-time and update recommended content; combined with natural language processing technology, it supports text/voice input to recognize user intent, match resources, and guide task completion.

Multi-source heterogeneous data integration

Multidimensional data, such as department dictionaries, doctor and professional information, and data from various subsystems, are integrated based on the knowledge base management system to assist the large model in accurately answering patients' questions during consultation.

04 Business Models

China Mobile's medical large model can provide comprehensive and differentiated solutions of "products + capabilities + services"

Products

Improve service efficiency through innovative information systems such as intelligent health assistants and full-cycle disease management, driving revenue growth in services like online consultation.

+

Capabilities

Provide medical large model capabilities, while promoting the development of new quality productivity in the industry through joint research and development innovation.

+

Services

Provide overall solutions and customized development services based on the needs of medical institutions.

05 Core Values

Improving patient satisfaction

In the pre-admission stage, the large model can extract key information such as the patient's condition and medical history with technologies like attention mechanism, reducing manual data entry. It automatically recommends departments for registration, and generates pre-consultation reports, preventing time and money losses caused by registration errors. It also utilizes the patient waiting time effectively to identify the focus in face-to-face consultation with doctors.

Providing authoritative access to specialized disease knowledge

It provides knowledge-based Q&A for specialized diseases based on such authoritative sources as clinical guidelines, medical textbooks and books, drug instructions, clinical pathways, and top journals. This helps patients obtain information about their conditions, medication guidelines, precautions, etc., so patients are less anxious and more emotionally stable.

Shorter waiting time

After patient registration and before face-to-face consultation, the multimodal large model can implement intelligent pre-consultation based on multi-source heterogeneous data integration and pre-training. The system can intelligently mine the patient's condition and medical history and automatically generate a pre-consultation report. The outpatient doctors can have a rough idea of patient condition during face-to-face consultation, improving consultation efficiency and reducing time on medical record writing. In a word, doctors can save at least 20% of their time on consultation and medical record writing.

Improving efficient, accurate diagnosis and treatment

In the full-cycle management of specialized diseases, the large model combines medical knowledge and expert experience to generate a chain of thought for clinical treatment based on CoT Prompting. The chain can support medical professionals in making decisions, developing treatment plans, and planning follow-up visits. Moreover, it will enhance treatment level and empower digital innovation in medical care.

Application Results

The First Affiliated Hospital of Fujian Medical University attaches great importance to digital transformation. Backed by technologies such as 5G and artificial intelligence, it has integrated intelligent capabilities within the hospital to enhance service capacity and treatment capability, and explored new smart service models to comprehensively improve the hospital's rating in this respect. With three months of launch, the project has served more than 100,000 patient visits. Thanks to it, the hospital has improved patient satisfaction and enhanced its image.

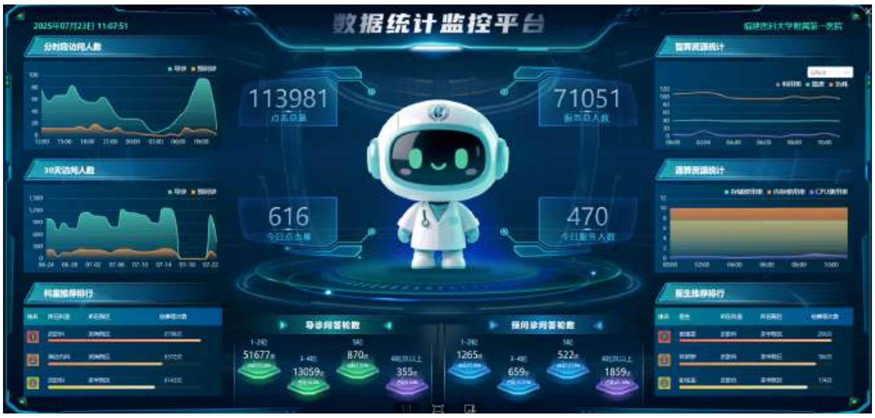


Figure 2: Console user statistics

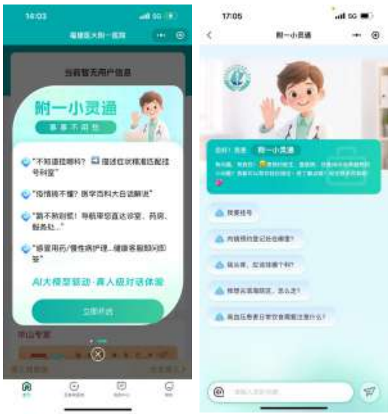


Figure 3: User interface

REFLECTION

Key Insights

By building smart hospitals based on large models with the First Affiliated Hospital of Fujian Medical University, China Mobile Fujian Company and Chengdu Institute have accumulated valuable experience. The deep integration of AI large models with hospital service processes has greatly improved patient satisfaction and reduced waste of patients' time and money. Moreover, it has optimized the disease management workflow through AI-driven full-cycle disease management, enhancing connections with patients.

Future Plans

With continuous progress and innovation in AI technology, we will continue to explore the application of large AI models in smart hospital construction, and expand the service scope to cover a wider range of emergency care scenarios and needs. We aim to build a smarter, safer, and more efficient hospital, making greater contributions to improving medical treatment levels and ensuring public health.

Innovative Practice of Large-Scale Personalized Education Utilizing Education-Specific Large Models

China's "Education Powerhouse Construction Plan (2024-2035)" outlines new demands for the deep integration of large language models into education and teaching. China Telecom has developed a large language model for the education industry based on high-quality teaching data, integrated with a precision teaching platform to provide intelligent lesson preparation, intelligent grading, and personalized tutoring. This empowers teachers to teach accurately and students to learn in a personalized manner. The solution is implemented in over 2,000 schools across China, promoting large-scale personalized education and supporting the development of a high-quality national education system.

Wu Jun
China Telecom Yizhi Education Technology Co., Ltd.
Vice President

SOLUTION PARTNERS

The large-scale personalized education solution powered by the education-specific large language model (LLM) focuses on improving students' overall subject competency under the new curriculum standards. Leveraging a high-quality dataset of over 100GB data consisting of test answers, grading data, courseware, assessment standards, and knowledge graphs, the solution employs an educational LLM trained via supervised fine-tuning (SFT) and reinforcement learning from human feedback (RLHF). It covers scenarios such as lesson preparation, intelligent evaluation of subjective questions, and tutoring. Integrated into a precision teaching platform that combines teaching, learning, practice, evaluation, and testing, and deeply embedded within classroom, homework, and examination scenarios, the solution reduces more than 70% of the teacher's workload in lesson preparation, assessment, and tutoring. By capturing full-scenario learning data from assignments and exams and applying big data analysis, it generates granular, knowledge point-based personalized learning reports for every student. This eliminates the need for repetitive problem-solving, boosting learning efficiency by at least 30%, and enables precise teaching and personalized learning at scale.

For example, Figure 1 illustrates the intelligent assessment of Chinese and English essays. The model provides not only scores but also detailed feedback and constructive suggestions for each student's essay, reducing the teachers' assessment workload, and helping improve students' language proficiency.

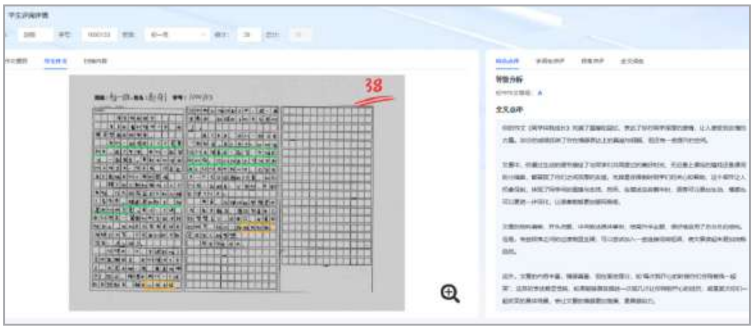
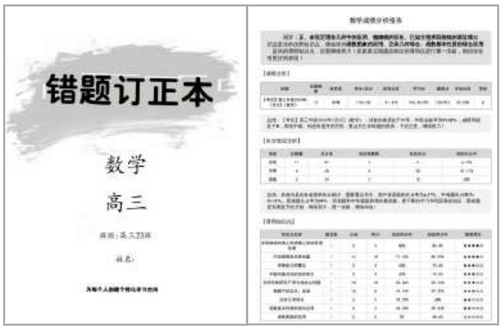


Figure 1: AI-powered scoring and multi-dimensional feedback for Chinese and English essays


错题订正本

数学

高三

Figure 2: Personalized learning report showing weak knowledge points andrecommending exercises for practice

The solution utilizes big data and LLMs to generate a personalized learning report for each student, detailed to the knowledge point level, as shown in Figure 2. This report features modules such as "performance analysis, difficulty analysis, knowledge point analysis, and targeted recommendations". By collecting assignment and exam data, it performs a comprehensive learning analysis, creating a customized workbook for each student. This approach allows students to focus on "incorrect" questions and solidify their understanding of correct answers.

Figure 3 shows the lesson plan generation LLM constructed from datasets of lesson plans and courseware. It facilitates collaborative lesson preparation across regions, reducing teachers' average lesson preparation time by 53%, increasing the reuse rate of courseware and resources to 78%, and minimizing redundant efforts. In terms of equitable resource distribution, the efficiency of obtaining quality lesson plans for county schools has improved by 400%, and the cross-school teaching and research participation rate has increased from 12% to 45%.

53 %	78 %	400 % ↑	45 % ↑ 12 %
reducing teachers' average lesson preparation time	increasing the reuse rate of courseware and resources	the efficiency of obtaining quality lesson plans for county schools	the cross-school teaching and research participation rate

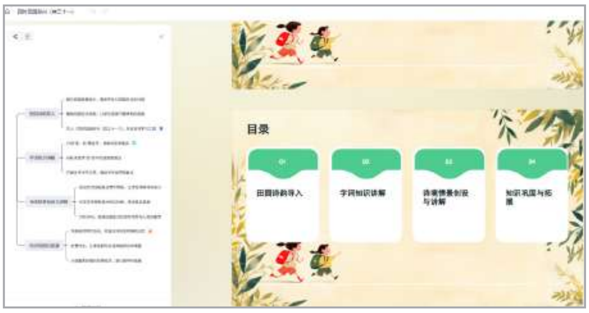


Figure 3: Teacher's lesson preparation assistant automatically generates courseware

In addition, to address the issue of students lacking immediate help with homework or self-study, the model simulates a Socratic dialogue, guiding students through answers step-by-step and providing support for difficult problems.

The "precision teaching platform" deeply integrated with the education-specific LLM is used in more than 2,000 schools serving 5 million teachers and students across 31 provinces, autonomous regions, and municipalities in China. It reduces teachers' grading workload by at least 70% and boosts the rate of excellent student writing by 60%. The entire system is developed with fully domestic and controllable technologies.

CHALLENGES

A key urgent challenge in education is leveraging AI to systematically reduce teachers' workload and boost students' learning efficiency, while avoiding negative impacts or ethical risks to teachers and students. From technical and customer needs perspectives, these challenges can be divided into the following five categories.

Eliminating the inherent hallucinations in general LLMs

Although general LLMs have excellent cross-task generation capabilities and human-like text interaction features, they still suffer from inherent "hallucinations". This appears as "authoritative-sounding fabrications" in areas like factual knowledge, numerical reasoning, and logical coherence. For example, committing a basic error like misjudging the numerical value "9.11" to be smaller than "1.19".

Lacking high-quality datasets and high annotation costs for building industry-specific LLMs

Existing public corpora are mainly composed of general text, lacking fine-grained annotations verified by both educational theory and cognitive science. Education-specific data is scattered among regional education research institutions, publishers, and examination centers. Expert annotation is extremely time-consuming (annotating a complete reasoning chain for a single secondary school math solution requires 15 to 30 minutes) and expensive.

Extremely low tolerance for value misalignment in the education industry, and low stability of existing LLMs in value-sensitive content

As a cornerstone of a nation, education must never contain content that violates its core values. However, responses to the same question across different dialog sessions or with minor prompt variations can produce outputs inconsistent with the Core Socialist Values. The lack of an explainable value conflict detection mechanism creates ethical risks for teachers and students.

Insufficient personalized analysis and inefficient interventions leading to resource misallocation

A common shortcoming in educational products is the inability to perform accurate learning analysis based on disciplinary ability frameworks, failing to systematically identify students' knowledge weaknesses, specific competency gaps, and literacy shortcomings. This forces parents and teachers to either manually review past exams and assignments (which is time-consuming), or enroll students in various tutoring classes (which is costly), trying to bridge learning gaps through "excessive exercises" and "repetitive tutoring."

Frequent, inefficient tasks limiting teachers' professional development opportunities

To improve teaching quality, teachers often increase the pace and frequency of assignments and exams within the limited class hours to identify individual and common learning issues. Consequently, they are burdened with vast amounts of routine work — preparing lessons, composing tests, grading, recording scores, analyzing statistics, and analyzing errors. This cycle of high-volume, low-value labor leads to burnout, depriving teachers of time for educational research.

SOLUTION AND VALUE

01 Overall Structure/Solution

The precision teaching solution powered by industry-specific LLMs is built around key technologies such as AI (large/small language models) and big data, addressing core educational scenarios of teaching, learning, practice, evaluation, and testing. Based on China Telecom's native cloud-network-edge security infrastructure, it builds an educational digital base, a big data center for personalized education, and an industry-specific LLM. By utilizing core AI capabilities such as educational knowledge graphs, intelligent grading, and learning path planning, it creates an integrated district-school application that features "intelligent lesson preparation, precision teaching, personalized learning, and district-school management." Through "monitoring and evaluation, problem analysis, target identification, precise teaching, and assessment and feedback," it enables precision teaching for teachers, personalized learning for students, and enhances overall regional teaching quality.

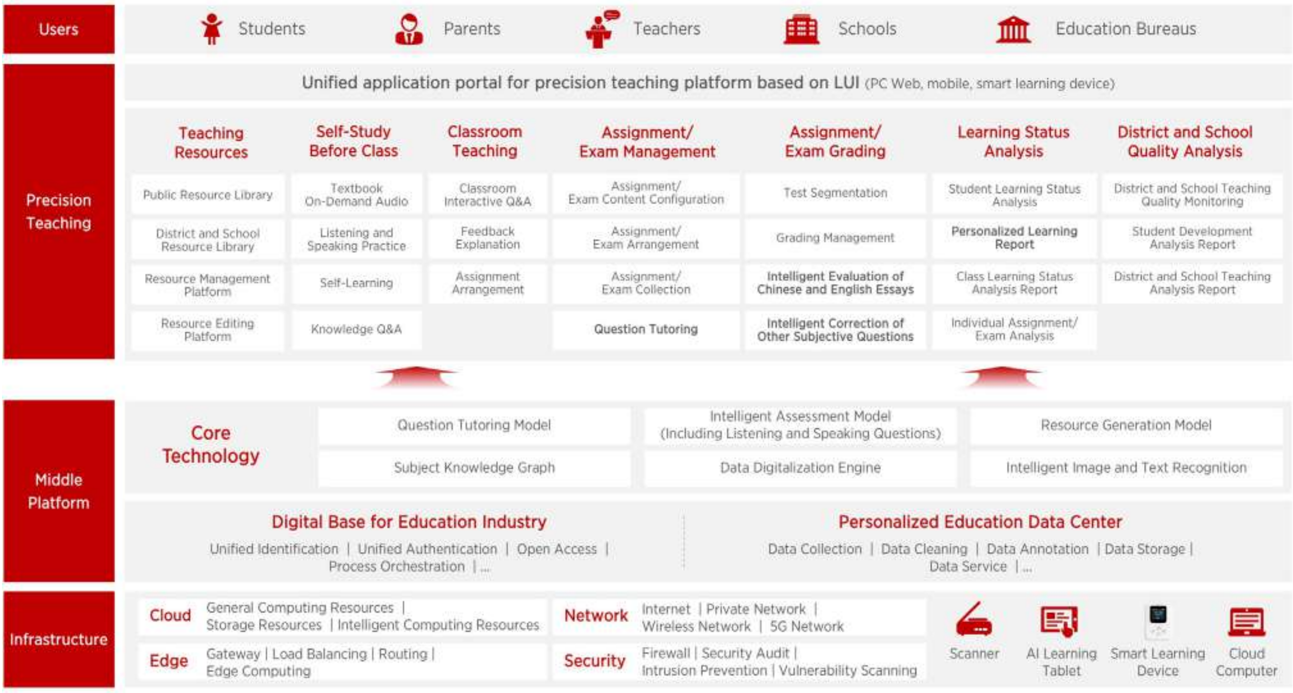

Users
Students
Parents
Teachers
Schools
Education Bureaus
Precision Teaching
Unified application portal for precision teaching platform based on LUI (PC Web, mobile, smart learning device)
Teaching Resources
Public Resource Library
District and School Resource Library
Resource Management Platform
Resource Editing Platform
Self-Study Before Class
Textbook On-Demand Audio
Listening and Speaking Practice
Self-Learning
Knowledge Q&A
Classroom Teaching
Classroom Interactive Q&A
Feedback Explanation
Assignment Arrangement
Assignment/ Exam Management
Assignment/ Exam Content Configuration
Assignment/ Exam Arrangement
Assignment/ Exam Collection
Question Tutoring
Assignment/ Exam Grading
Test Segmentation
Grading Management
Intelligent Evaluation of Chinese and English Essays
Intelligent Correction of Other Subjective Questions
Learning Status Analysis
Student Learning Status Analysis
Personalized Learning Report
Class Learning Status Analysis Report
Individual Assignment/ Exam Analysis
District and School Quality Analysis
District and School Teaching Quality Monitoring
Student Development Analysis Report
District and School Teaching Analysis Report
Middle Platform
Core Technology
Question Tutoring Model
Intelligent Assessment Model (Including Listening and Speaking Questions)
Resource Generation Model
Subject Knowledge Graph
Data Digitalization Engine
Intelligent Image and Text Recognition
Digital Base for Education Industry
Unified Identification | Unified Authentication | Open Access | Process Orchestration | ...
Personalized Education Data Center
Data Collection | Data Cleaning | Data Annotation | Data Storage | Data Service | ...
Infrastructure
Cloud
General Computing Resources | Storage Resources | Intelligent Computing Resources
Edge
Gateway | Load Balancing | Routing | Edge Computing
Network
Internet | Private Network | Wireless Network | 5G Network
Security
Firewall | Security Audit | Intrusion Prevention | Vulnerability Scanning
Scanner
AI Learning Tablet
Smart Learning Device
Cloud Computer

Figure 4: Precision teaching project architecture

02 Application Scenarios

Assignment and assessment management for district education authorities, education groups, and individual schools

Facilitates the data collection, intelligent assessment, personalized learning analysis, and regional teaching quality analysis for assignments and examinations in the districts, education groups, or individual schools.

Lesson preparation for elementary and secondary school teachers

Helps teachers define teaching objectives, analyze teaching materials, integrate resources, design processes, and predict challenging areas, achieving precision teaching and efficient classroom management through systematic lesson preparation.

Self-learning guidance for students/parents

Offers guidance for students' homework or practices.

03 Key Innovations

Ensuring industry-leading model accuracy through data feedback and reinforcement learning from human feedback (RLHF)

Initially, the intelligent assessment model for Chinese and English essays was trained using supervised fine-tuning (SFT) based on expert-annotated data, and the model concordance can reach over 88%. However, teachers still need to make corrections for high scores, low scores, and possible anomalies, which not only increases the teachers' workload but may also affect student engagement due to scoring errors. By desensitizing and feeding back the assessment data from teachers during the school deployment process, applying it in RLHF, and enhancing with retrieval-augmented generation (RAG) for outliers, the LLM's concordance has been improved to over 95%, setting an industry benchmark. This approach also allows the model to uphold uniform assessment criteria while accommodating the grading strictness of different schools through data fitting.

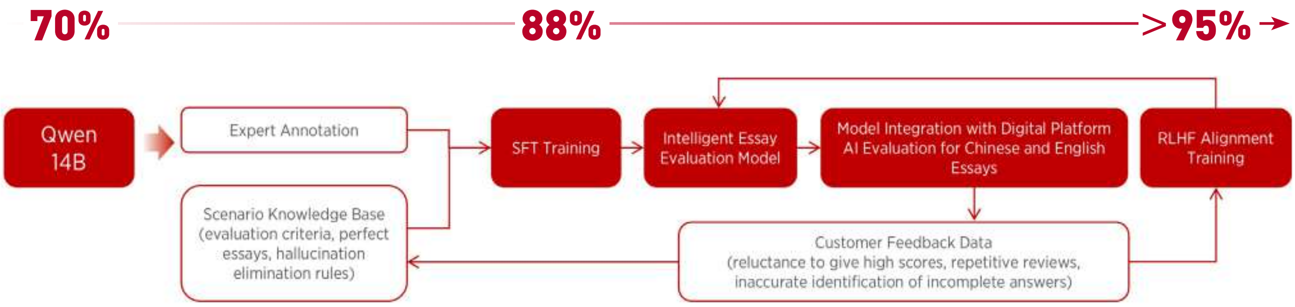

70%
88%
>95%
Qwen 14B
Expert Annotation
Scenario Knowledge Base (evaluation criteria, perfect essays, hallucination elimination rules)
SFT Training
Intelligent Essay Evaluation Model
Model Integration with Digital Platform AI Evaluation for Chinese and English Essays
RLHF Alignment Training
Customer Feedback Data (reluctance to give high scores, repetitive reviews, inaccurate identification of incomplete answers)

Figure 5: Model training process

AI self-annotation addressing high annotation costs

In large-scale data annotation for intelligent subjective question assessment, using expert-defined dimensions and samples still faces challenges of high cost and unstable data quality. Implementing AI data annotation, where the LLM performs initial annotation on raw data followed by expert correction, ensures data quality for SFT training. Methods like cross-checking and comparative quality analysis between AI and benchmark expert annotations are used.

Enhanced core value alignment

During LLM training, in addition to incorporating the latest core value corpus in SFT training data, the Aligner technology is integrated with RLHF. This adds real-time screening and correction functions, interpreting and understanding the semantics of model inputs and outputs to guarantee adherence to educational compliance and security requirements. For content in minor education, Aligner can automatically filter and correct inappropriate outputs, ensuring the appropriateness and security of educational materials.

04 Business Models

The precision teaching platform powered by the education-specific LLM follows China Telecom's philosophy of intelligent benefit. Capitalizing on its natural government, business, household, and customer (GBCH) alignment, the overall marketing strategy is "drive volume via widespread adoption in government and business organizations, and generate revenue from value-added services for household and individual customers."

For education authorities (government)	The focus is on integrating district and county platforms with intelligent education platforms previously established by China Telecom for provincial, city, district education bureaus, typically through annual service purchases. This approach builds comprehensive regional systems for intelligent teaching, examination and assessment (including joint examinations), and assignments, serving teachers and students in schools across the province, district, and county. Examples include the Ningxia Hui Autonomous Region Smart Assignment Project and the Jiangxi Province Smart Education App Project.
For individual schools (business)	Services are provided for regular school activities such as weekly and monthly tests, assignments, and lesson preparation, with an average annual service fee of 30,000 yuan.
Value-added services for parents and students (household and customer)	By analyzing student data from assignments, exams, and independent learning, the system generates personalized learning analysis reports. Parents can purchase monthly value-added plans to obtain these reports or access tutoring services.
Integrated hardware and software	The precision teaching platform and tutoring services are integrated into AI learning tablets sold to consumers.

05 Core Values (including quantitative results and qualitative impact)

The precision teaching platform powered by the industry-specific LLM reduces teacher workload through AI, giving them more time for educational research and teaching quality enhancement. It also provides students with equitable access to high-quality tutoring, supporting parents who lack the time or capacity for direct academic guidance.

At the same time, the platform prioritizes supply chain security and alignment with core values, maintaining control over critical technologies and stable supply of key resources. It ensures all input content complies with regulations and protects students' well-being.

In practical applications, teachers' examination grading efficiency has increased by 60%, cutting per-paper processing time from 7 minutes to 2-3 minutes. For a teacher responsible for 3 classes (around 135 students), this saves roughly 6 hours of manual grading daily. It covers 100% of subjects, overcoming the time-consuming challenge of grading subjective questions in traditional teaching. For students, utilizing Socratic questioning and personalized error analysis boosts answer accuracy by 20% over time. Periodic testing indicates a 15% increase in students' knowledge mastery compared to before.

Regarding business impact

- **2022-2023**
  The average annual revenue was 40 million yuan in 2022-2023
- **2024**
  After the precision teaching platform was improved with LLMs, the average annual revenue in 2024 grew to 80 million yuan
- **2025**
  In 2025, the plan is to expand to 3,000 schools, with an estimated revenue of 100 million yuan, potentially driving a further 200 million yuan in associated telecom and cloud network revenue.

Figure 6: Proof from actual school deployment

REFLECTION

Key Insights

Key lessons from developing and scaling the industry LLM-powered precision teaching platform for personalized education include:

- **Industry LLMs excel in domain accuracy:** For high-precision scenarios like test grading and tutoring, the industry model significantly outperforms general models (such as GPT-4, full-featured DeepSeek). General models, due to insufficient domain-specific training data, struggle to match professional teacher expertise.
- **Product-model integration is crucial:** Teachers and students need a platform deeply integrated with the education system, not an isolated LLM tool. Through "product-model integration," models are embedded into digital platforms to reduce the usage friction.
- **RLHF and RAG effectively limit hallucinations:** The application of RLHF and RAG can significantly cut the model's hallucination rate from 10% to under 1%, addressing problems including scoring errors and output repetition.

Future Plans

Future platform optimization will focus on the following three aspects based on customer feedback.

- **Large and small model collaboration:** An mixture-of-experts (MoE) structure will enable dynamic routing among the core industry LLM, specialized vertical scenario models, and small models (like OCR/computer vision), facilitating flexible cooperation and an optimal accuracy-efficiency trade-off.
- **Agent ecosystem:** Developing a matrix of intelligent agents for core functions—teaching, evaluation, research, and management—using model context protocol (MCP)/agent-to-agent (A2A) standards, creating an adaptive ecosystem of educational agents.
- **Inference cost reduction:** A triple-focus approach will target computational architecture (such as leveraging cascaded low-cost inference GPUs), algorithms (employing 8-bit quantization, speculative decoding), and API services (implementing high-concurrency asynchronous text output) to exponentially lower inference costs, making the technology more affordable for teachers and students.

Civil Aviation Customer Service Large Model and Intelligent Applications

> **Against the backdrop of AI technology profoundly reshaping industry ecosystems, the civil aviation customer service sector faces dual challenges of upgrading both service efficiency and user experience. Leveraging the collaborative innovation mechanism of central state-owned enterprises (SOEs), China Mobile and TravelSky jointly tackled these challenges to create the "Civil Aviation Customer Service Large Model and Intelligent Applications" project. Through the deep integration of 5G, large models, and civil aviation business systems, they built the industry's first end-to-end intelligent customer service solution, opening a new chapter in the construction of smart civil aviation.**
>
> **Feng Junlan**
> **Chief Scientist of China Mobile Group /**
> **China Mobile Jiutian Artificial Intelligence Technology (Beijing) Co., Ltd.**
> **(Jiutian Artificial Intelligence Research Institute)**

SOLUTION PARTNERS 中国移动 China Mobile | 中国航信 TravelSky

With the recovery growth in civil aviation passenger traffic, residents' travel and production transport demands are rapidly increasing, becoming more diverse and personalized, leading to higher demands on the quality and efficiency of civil aviation customer service. Traditional customer service models reveal numerous issues, such as slow response from human agents, limited service hours, and inefficiency in handling large volumes of repetitive inquiries, making it difficult to meet passenger expectations. The intelligent transformation and upgrade of the civil aviation customer service is thus extremely urgent.

In 2023, the Civil Aviation Administration of China released the "Smart Civil Aviation Construction Roadmap," explicitly proposing the requirement to "promote the large-scale application of AI in customer service," incorporating intelligent customer service into the core construction indicators of Smart Civil Aviation. The State-Owned Assets Supervision and Administration Commission's (SASAC) "Notice on Accelerating the Digital Transformation of State-Owned Enterprises" emphasized that central SOEs need to become role models in fields like AI and 5G, providing policy backing for the intelligent transformation of the civil aviation industry.

Driven by industry demands and national policy, China Mobile actively responded, collaborating with TravelSky to establish a joint research project in 2024, exploring the application of large models, AI agents, and other AI technologies in civil aviation customer service. Based on the Jiutian Zhongqing Base Large Model (75 billion parameters), and integrating 68,000 supervised fine-tuning (SFT) civil aviation data points and 5GB of industry documents, they co-built the civil aviation-oriented "Qihang" large model and the industry's first large model-based civil aviation customer service product – AirVoice Intelligent Customer Service. This project created six types of benchmark applications, including 5G video call, intelligent messaging, agent assist, and ticket automation, reducing average agent interaction time by over 50%, increasing the diversion rate of manual calls by over 20%, and achieving a full-channel service closed loop. It has currently served over 30,000 business users, becoming a benchmark case for Smart Civil Aviation construction.

50%+
reducing average agent interaction time

20%+
increasing the diversion rate of manual calls

CHALLENGES

Manual efficiency and resource misallocation: A large number of inquiries are repetitive yet overly reliant on manual processing, leading to high labor costs. During specific peak periods, passenger wait times could exceed 15 minutes, creating a "high cost, low efficiency" resource dilemma that increases operational burden and tries user patience.

Insufficient dynamic adaptation of knowledge systems: Civil aviation policies, flight statuses, and other information change frequently, but traditional knowledge bases are typically updated weekly. Simultaneously, service personnel require additional time for training, resulting in low answer accuracy, severely impacting service professionalism and user trust, and forming a vicious cycle of "information lag - incorrect responses - increased complaints."

Aging interaction channels and experience disconnect: Current customer service operations rely heavily on telephone channels, lacking new interaction methods like 5G video calls and 5G intelligent messaging. This disconnect from the digital experiences preferred by younger users, coupled with channel singularity, restricts the breadth and depth of service reach.

High enterprise management costs: As the primary point of contact for passengers, enterprises must invest significant costs in training customer service personnel, with cultivating a senior agent requiring substantial human and time resources. After sessions, agents also need considerable time to fill out session tickets, reducing customer-facing frequency. Furthermore, given vast amounts of customer service conversations, only manual sampling of session content is possible, preventing comprehensive and effective monitoring of response standardization.

SOLUTION AND VALUE

01 Overall Structure/Solution

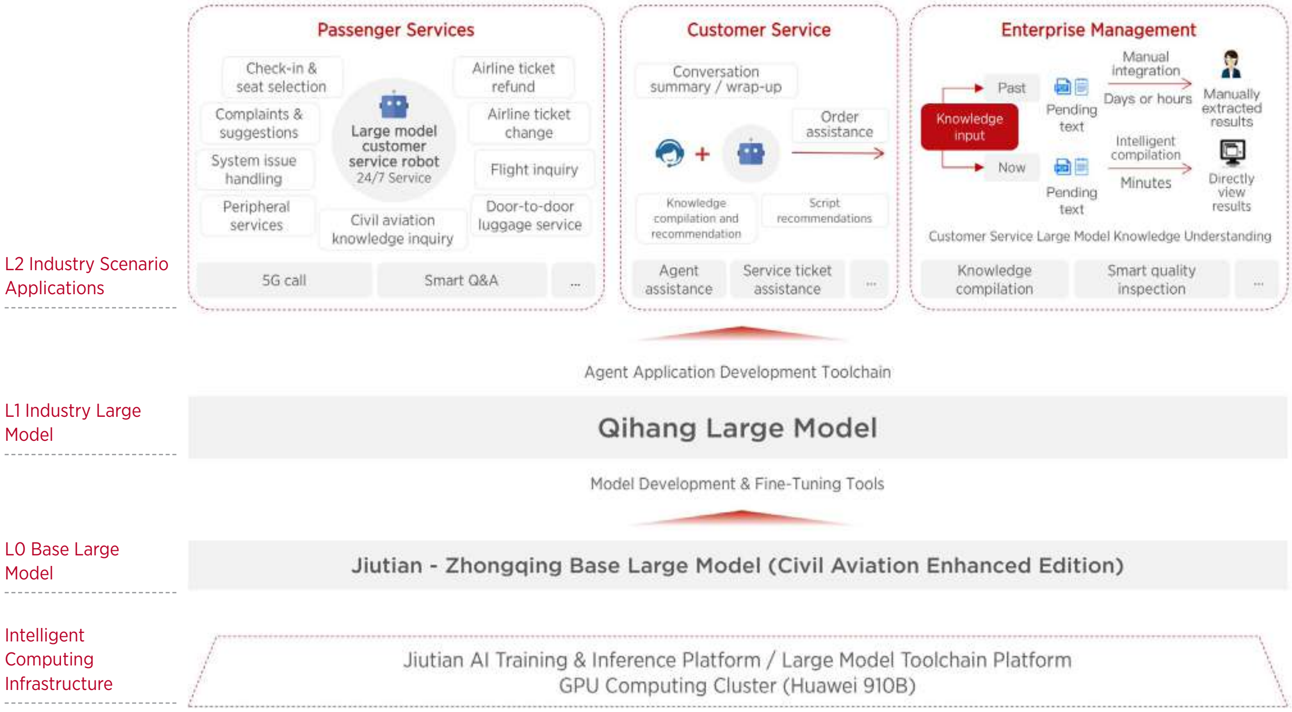

L2 Industry Scenario Applications
Passenger Services
Check-in & seat selection
Airline ticket refund
Complaints & suggestions
Large model customer service robot 24/7 Service
Airline ticket change
System issue handling
Flight inquiry
Peripheral services
Civil aviation knowledge inquiry
Door-to-door luggage service
5G call
Smart Q&A
...
Customer Service
Conversation summary / wrap-up
Order assistance
Knowledge compilation and recommendation
Script recommendations
Agent assistance
Service ticket assistance
...
Enterprise Management
Knowledge input
Past
Now
Pending text
Pending text
Manual integration
Days or hours
Manually extracted results
Intelligent compilation
Minutes
Directly view results
Customer Service Large Model Knowledge Understanding
Knowledge compilation
Smart quality inspection
...
Agent Application Development Toolchain
L1 Industry Large Model
Qihang Large Model
Model Development & Fine-Tuning Tools
L0 Base Large Model
Jiutian - Zhongqing Base Large Model (Civil Aviation Enhanced Edition)
Intelligent Computing Infrastructure
Jiutian AI Training & Inference Platform / Large Model Toolchain Platform
GPU Computing Cluster (Huawei 910B)

Figure 1: Overall Architecture

Intelligent computing infrastructure layer

Based on the Jiutian AI Training & Inference Platform, it manages diverse AI computing power and deep learning frameworks, providing full-stack, domestically-controlled, full-lifecycle large model construction services from design, training, fine-tuning, compression, deployment, hosting, to inference.

L0 base layer

Based on the Jiutian Zhongqing Base Large Model, it incorporates extensive general knowledge of the civil aviation domain to build an enhanced model for civil aviation.

L1 industry layer

Leveraging the L0 base model capabilities, it uses 68,000 professional Q&A SFT data points and 5GB of industry-specific documents, fine-tuned via the Jiutian AI Platform, to form the professional "Qihang" large model tailored for civil aviation customer service.

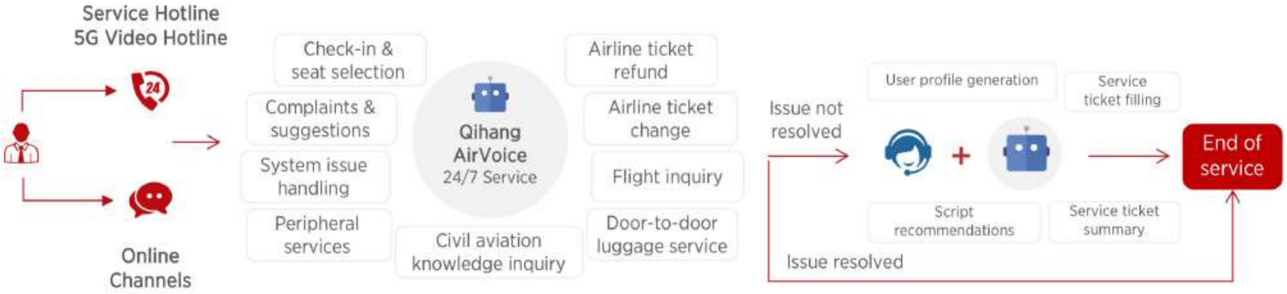
L2 scenario application layer

Based on the L1 Qihang model, utilizing the rich tool capabilities of the agent platform and advanced agent construction technologies like Langchain and Workflow, it establishes a closed-loop data flywheel mechanism ("business data feedback - automated labeling - model optimization - capability evaluation"). This achieves a 50% increase in data utilization and an 8-fold improvement in labeling efficiency, enabling efficient benchmark applications like 5G video customer service, intelligent messaging, agent assist, and knowledge compilation.

02 Application Scenarios

Identity Verification
Service Confirmation

Uses methods like facial recognition/comparison and online signatures to optimize the processes of passenger identity verification and service procedure confirmation, making business process handling safer and smoother.


Service Hotline
5G Video Hotline
Online Channels
Check-in & seat selection
Complaints & suggestions
System issue handling
Peripheral services
Qihang AirVoice 24/7 Service
Civil aviation knowledge inquiry
Airline ticket refund
Airline ticket change
Flight inquiry
Door-to-door luggage service
Issue not resolved
User profile generation
Service ticket filling
Script recommendations
Service ticket summary
End of service
Issue resolved

Omni-channel intelligent customer service

5G video customer service

Supports functions like facial recognition and online signature, improving identity verification efficiency by 70%.

Intelligent message interaction

Through image/text recognition and multi-turn dialog, it fuels a self-service rate of over 80% for services like flight inquiries and refunds/changes.

Agent efficiency enhancement

Precise knowledge recommendation

Real-time pushing of business scripts and operational guidance during calls reduce agent search time by 60%.

Automated ticket generation

Automatically generates tickets based on dialog semantics, reducing processing time from 4 hours to 20 minutes.

Intelligent quality inspection & operations

Full-scope voice quality inspection

Covers 100% of call recordings, with violation/non-compliant speech recognition accuracy exceeding 95%.

Intelligent knowledge base management

Document compilation time is reduced from 2 hours to 10 minutes, and knowledge update cycle shortened to T+1.

03 Key Innovations

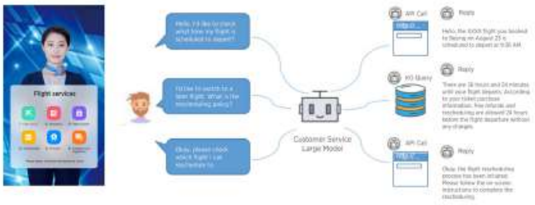
Multi-modal integrated interaction	Supports mixed input of voice, text, and images, with a recognition accuracy rate of 85%, solving consultation challenges related to graphical interface operations.
Deep infusion of industry knowledge	Integrates data from multiple systems, including airlines, airports, and agents, to construct a proprietary knowledge base containing civil aviation regulations, passenger rules, and professional terminology. Trained on over 100,000 hours of real customer service voice dialogs, it accurately parses complex scenarios like "through check-in baggage" and "mixed cabin class changes."
End-to-end ecological sustainability	Based on NPU computing power, a complete technical chain from model training to inference was built, achieving full-stack sustainability from chips and frameworks to agent applications.

04 Business Models

Central SOE collaboration model	Establishes a "Five Shared" cooperation mechanism ("shared development, shared investment of resources, shared risks, shared results, shared benefits"), sharing intellectual property and benefits, reducing R&D risks for individual enterprises.
Tiered service system	Basic edition: Provides SaaS-based intelligent customer service for small and medium-sized airlines, with an annual fee model that lowers the barrier to entry. Custom edition: Offers localized deployment solutions for large airports, charging based on the number of agents and service volume.
Ecological win-win	Opens API interfaces to collaborate with other civil aviation apps and platforms in building a civil aviation service ecosystem.

05 Core Values (including quantitative results and qualitative impact)

Enhancing the Service Experience for Air Travelers



Large model customer service robot

Before LLM

- Most service procedures required manual intervention, which was inconvenient
- Long wait times for an agent, leading to a poor experience
- Cumbersome identity verification methods during customer service; unfriendly service processes

After LLM

- Online self-service
- AI-powered customer service for instant responses
- Diverse and fast identity verification that significantly enhances customer experience

Improving Service Efficiency for Airline Agents

Knowledge recommendation

Conversation summary / wrap-up

Service ticket assistance

Knowledge base assistant

Large model agent assistance

Before LLM

- Laborious and time-consuming service ticket entry and summarization
- Non-standardized service scripts
- Inability to provide personalized scripts and solutions based on the customer

After LLM

- Significantly enhanced customer service reception capacity
- Greatly improved service ticket processing efficiency
- Personalized customer service based onconversation summaries

Quantitative results

Efficiency gains

40% Ticket processing time reduced

60% document analysis time decreased

30% agent involvement rate lowered

Experience upgrade

95%+ Recognition accuracy for high-frequency business scenarios in intelligent customer service reaches

90%+ medium-low frequency scenarios

93% overall intent recognition accuracy

Impact

1 large model in civil aviation

9 Fills a domestic gap, becoming the 9th large model filed/recorded among central SOEs supervised by SASAC

Forms a replicable solution of

Large Model + **Robots** + **Business System**

expected to cover multiple airports and airlines by 2026.

REFLECTION

Key Insights

During the critical period of civil aviation digital transformation, China Mobile and TravelSky, guided by national strategic needs, leveraged the dual-drive model of "Technology + Industry" and the advantages of central SOE collaboration to create a full-chain innovation ecosystem from "basic research to technological breakthrough and scenario implementation." Using the "Qihang" large model as the technical foundation, they deeply integrated computing power advantages with civil aviation scenarios, developing product matrices like intelligent customer service, forming a closed-loop mechanism from demand insight to value verification. Through innovations like data sandboxes and adaptation to domestic computing power, they built a self-controlled and secure technological system, promoting the leap of civil aviation services from standardized supply to precise matching. This provides a multidimensional example of "technology penetrating scenarios, scenarios nurturing technology" for cross-domain collaboration among central SOEs, establishing a benchmark for central SOE collaborative innovation.

On the product-driven front, with the "Qihang" large model as the core engine, they developed innovative products like the Qihang Intelligent Customer Service AirVoice, forming a closed loop from technology R&D to business implementation. Considering practical business needs such as large model hallucinations and business users' stringent requirements for answer consistency, they innovatively designed the closed-loop data flywheel mechanism of "business data feedback - automated labeling - model optimization - capability evaluation - business optimization application," breaking through the efficiency bottleneck of traditional manual data labeling.

Future Plans

Next, China Mobile and TravelSky will continue their in-depth cooperation, leveraging mechanisms like the Central SOE Innovation Collaboration Platform. While further enhancing the user experience of commercial products, they will actively expand external airport and airline customers, creating a new model for tripartite cooperation. Building on this, they will unite industry, academia, and research to create a joint technology innovation platform involving multiple participants, providing dual empowerment through technology and business for the digital-intelligent transformation of the civil aviation field.

Harnessing AI to Reshape Enterprise Training Efficiency

> As AI and large language models (LLMs) advance, the need for intelligent enterprise training grows more pressing. Traditional training methods struggle to meet the goals of high efficiency, accuracy, and personalization. China Telecom's Xingchen Agent Platform serves as an all-in-one platform for agent creation and management. It enables code-free development of AI-enhanced education applications, featuring optimization abilities and industry-level security. This creates a closed-loop iterative process that solves LLM application difficulties, helping to transform enterprise training productivity.
>
> **Zhongjiang He**
> **China Telecom Artificial Intelligence Technology (Beijing) Co., Ltd.**
> **Secretary of the Party Committee, Chairman of the Board,and CEO of China Telecom Artificial Intelligence Technology**

SOLUTION
PARTNERS

In the tide of the digital economy and intelligent transformation, and grounded in the demand for high-quality enterprise development, China Telecom has introduced the "AI Search" knowledge service application. This offering is built upon the Xingchen Agent Platform, featuring deep integration of the DeepSeek LLM and enterprise knowledge base. It provides a new smart dynamic for cultivating enterprise talent and training management personnel. This innovation tackles the shortcomings of traditional training methods, establishes a new industry standard for digital transformation, and offers powerful support for talent strategies in the modern enterprise landscape.

CHALLENGES

Facing issues like rapidly updating enterprise knowledge, increasingly varied staff learning requirements, and fragmented training resources, traditional training approaches exhibit several significant limitations:

Delayed knowledge updates

According to Deloitte's "2024 Global Human Capital Trends" report, major shifts in the knowledge and skill frameworks of sectors such as technology and finance occur about every 18 months. However, research by Training Industry shows that more than 70% of enterprise training courses are updated less frequently than every 12 months. Some manufacturing companies undertake a full update only once every 3 to 5 years. Consequently, the knowledge employees gain from training becomes misaligned with practical job requirements. For example, during periods of accelerated digital transformation, the understanding of digital tools that employees receive is frequently outdated, unable to support the company's operational expansion needs.

Challenges in accurately addressing varied learning needs

According to LinkedIn's "2024 Workplace Learning Report", while 89% of employees worldwide recognize the importance of personalized learning for their professional growth, only 23% believe company training adequately addresses their personal needs. Furthermore, a Gallup survey demonstrates that in the traditional uniform training approach, 63% of staff perceive the training content as having low relevance to their jobs and being non-specific, which directly undermines engagement and substantially compromises the training results.

Inefficient resource discovery and challenges in assessing training results

Studies from Accenture find that employees typically require 1 to 3 hours to locate appropriate learning resources in a traditional training setup, severely affecting the efficiency and motivation for independent learning. Additionally, Brandon Hall Group's report highlights that more than half of all companies struggle with assessing training results. The absence of robust data collection and analysis tools impedes accurate measurement of how training translates into improved employee performance, making it hard to refine training strategies effectively.

The problems outlined above create significant obstacles for monitoring and assessing training efficacy, consequently preventing training programs from reaching their fundamental goals of being highly efficient, precise, and personalized.

SOLUTION AND VALUE

01 Overall Structure/Solution

With "AI Search" at its core, China Telecom established a three-layer technical framework of "knowledge processing, low-code development, and dynamic model adaptation" to develop a smart training system that is lightweight, adaptable, and context-aware. The platform uses the Xingchen Agent Platform's retrieval-augmented generation (RAG) technology to consolidate data from multiple sources, automatically analyzing unstructured content such as training courses and industry reports for immediate knowledge retrieval and application. Low-code workflow design enables quick responses to evolving business demands. A dynamic model adaptation mechanism allows intelligent allocation of computing power, delivering tailored learning paths for various job positions and significantly boosting training effectiveness.

02 Application Scenarios

Enterprise training	Education scenarios
During new employee onboarding, "AI Search" rapidly consolidates information on company culture and operational procedures specific to their role, providing personalized learning paths to facilitate swift integration. Technical staff can obtain up-to-date technical documentation and case analyses, keeping their skills current.	In vocational training, for disciplines like e-commerce, "AI Search" combines real-time industry updates and practical skills to tailor hands-on learning programs for students. In standard K-12 education, it helps teachers swiftly locate appropriate teaching resources and customize lesson plans based on student needs, while also offering students personalized supplementary content.

03 Key Innovations

"AI Search," built upon the Xingchen Agent Platform, combines LLMs with knowledge bases to automatically analyze and synthesize resources, dynamically fitting needs and enabling immediate knowledge retrieval and application, which greatly increases both efficiency and precision. In terms of technical innovation, the Xingchen Agent provides "AI Search" with end-to-end optimization. Through collaboration with the Xingchen LLM, it carries out targeted enhancements for agent scenarios on preset models and tools, resulting in more than a 30% boost in end-to-end performance. A data feedback and fine-tuning mechanism also enables the agent to self-optimize continuously. In addition, the platform is equipped with independently developed safety fence safeguards, satisfying the stringent security requirements of state-owned and central enterprises, and delivers refined permissions management for knowledge bases, toolkits, and agent applications. These technical capabilities drive substantive upgrades in training and educational efficacy, presenting a replicable practical blueprint for smart transformation across industries.

Precise retrieval and efficient knowledge access	Through vector search and semantic analysis techniques, the platform converts massive amounts of enterprise knowledge—including internal regulations, industry standards, and strategic documents—into structured assets. This supports managers and staff in rapidly accessing critical information such as the latest industry updates and operational guidelines, aiding scientific decision-making.
Dynamic adaptation and flexible scenario coverage	Lightweight models guarantee immediate response during high-traffic periods, whereas LLMs provide deep interpretation of complex business problems (for instance, optimizing market strategies or formulating technical plans). This allows for flexibility across all scenarios, from fundamental skill training to high-level management discussions.
Closed-loop optimization for continuous service improvement	A cycle of "user feedback, model refinement, and knowledge base updating" is implemented to correct knowledge inaccuracies promptly and keep industry cases current. This maintains alignment between training materials and corporate strategy, ensuring learning achievements are precisely converted into operational competence.

04 Business Models

The commercial development of the Xingchen Agent is based on creating standard products, training frameworks, and service methods through comprehensive internal usage across the group. Externally, it constructs a collaborative ecosystem where products, training, and services are key elements. Internally, the standardized Group Edition is mandated for use group-wide, helping to accumulate technical expertise and establish uniform demand response standards. For external developers and enterprise/government customers, open-source and public cloud versions are available to suit varying deployment and security needs, thereby closing the loop from "capability reuse" to "market expansion" and "mutual benefits."

Group-level platform and comprehensive operational services for internal group and large state-owned enterprises

This model prioritizes "cost efficiency." The Group Edition eliminates redundant development costs, serving as the central agent platform for diverse internal business needs, providing custom solutions, and enhancing external versions with refined, common capabilities.

Flexible open source platform for technical users

Centered on "adaptability," this model invites provincial companies and partners to co-build, increasing industry influence. It delivers core code frameworks and standard APIs to provincial companies and government and enterprise customers, enabling secondary development aligned with local operations (such as area-specific training). After private deployment, remote technical assistance is provided by China Telecom. Revenue is generated through a basic license fee, plus pay-for-service customization (such as custom module development, local operation support, expert consultation).

Public cloud product model for individual and small enterprise developers

Leveraging the Xirang all-in-one intelligent platform, it provides SaaS services to external business customers. The service enables online registration for immediate, out-of-the-box operation, catering primarily to small and medium enterprises. A subscription-based payment model (annual or monthly plans) is available.

A key point of flexibility is the range of private delivery options provided by the Xingchen Agent.

Lightweight application delivery

The platform allows for standalone delivery of specific agent applications (excluding the platform itself). This enables enterprises to deploy pre-built agents (such as panoramic sensing agent and customer service agent) directly without constructing a complex technical structure. It is tailored for businesses with limited technical resources that prefer an "out-of-the-box" experience, simplifying accessibility.

Independent platform building

The agent platform can be delivered separately. Enterprises gain access to a full suite of tools (for model training, application orchestration, and data management), allowing them to build and refine agents tailored to their operations on the Xingchen platform. This caters to users with dedicated IT resources pursuing in-depth customization, facilitating the creation of a unique AI ecosystem.

Full-stack solution delivery

An integrated "agent platform and application" solution can be delivered. This includes a scalable platform base supplemented with ready-made agents designed for specific industry contexts (such as long-text writing agent, credit approval agent, and HR hiring assistant). Enterprises can use these proven applications immediately or customize them via the platform, addressing both efficiency and customization requirements, and providing end-to-end support from basic use to advanced operation.

Through layered version design and collaborative operations, it ensures internal capability reuse and efficiency improvement for China Telecom while expanding the external market with differentiated services, building a sustainable cycle of commercial value.

05 Core Values

"AI Search" has served more than 700,000 users. Since its launch, it has assisted in solving over ten thousand frontline problems and boosted the efficiency of knowledge retrieval by over 60%, becoming a vital "intelligent assistant" for management training and staff skill improvement. According to feedback from a training supervisor at a branch company: "Using 'AI Search,' new employees can rapidly grasp operational procedures, while managers can obtain the latest industry insights. This has reduced the training duration by 30% and significantly increased our team's overall efficiency." This success showcases China Telecom's innovation strength and demonstrates the pioneering role of state-owned enterprises in driving the digital transformation of talent management.

REFLECTION

Key Insights

From a technology application perspective, the Xingchen Agent Platform has accurately met diverse customer needs in various situations, leveraging powerful AI capabilities like multi-modal interaction and smart intention comprehension. Practical applications, spanning from automating business workflows to improving learning outcomes in educational contexts, have demonstrated that technology serves as the fundamental force behind deploying intelligent services, supporting effective demand fulfillment and problem-solving.

In terms of cooperation models, working closely with customers is essential. By deeply understanding customer requirements and conducting industry-specific custom development, we guarantee that the agent applications can effectively address customers' pain points. Additionally, a professional training and support framework offers ongoing reliability, enabling customers to rapidly adopt and utilize agents.

Future Plans

In the future, we will continue to increase investment in technology research and development, continuously optimize the features and performance of the Xingchen Agent Platform, and enhance its adaptability and intelligence level in more complex scenarios. The roadmap involves launching super agents, additional MCP tools, agents for specific verticals, diversified front-end capabilities for agents, and advanced multi-agent coordination. These will serve domains like government affairs, banking, education, public security, and transportation, reinforcing interoperability across different contexts. Agent self-optimization will be achieved via assessment tools and model optimization, while open APIs will foster a partner-driven technology ecosystem for faster adoption.

For market expansion, the strategy is to broaden reach into more industries, creating standardized solutions addressing common industry-specific requirements. Successful implementations will serve as references to attract more customers. We will also intensify co-innovation with ecosystem partners to promote agent utilization across industries, playing a greater role in supporting businesses' digital transformation and intelligent upgrade.

"Smart Clothing": AI-Driven Digital Relics and Immersive Cultural Experiences

> In response to digital transformation needs of the cultural creative and tourism industry plagued by such issues as singular forms of communication, weak interactive experiences, and homogenized creative products, China Unicom has launched the "Smart Clothing" interactive experience platform based on its multimodal large model "UniAI" (Yuanjing) and industry data. The platform presents cultural institutions like museums and scenic spots with new ways of display and operation, bringing immersive, personalized experiences to young users. By reviving cultural heritage through AI technology, the platform enhances visitor engagement and commercial benefits and promotes cultural inheritance and industry upgrade. The vision of "making history tangible, culture accessible, and heritage sustainable" is realized.

Li Zhenjun
China Unicom Data Intelligence Co., Ltd.
Deputy General Manager

SOLUTION PARTNERS

In response to the surging demand for immersive, personalized cultural experiences and the national digital cultural strategy, China Unicom has launched the "Smart Clothing" interactive experience platform with its independently developed multimodal large model "UniAI" and extensive data accumulated in the museum industry.

The platform targets two major groups: For cultural operators (museums, scenic spots, cultural companies), it solves challenges in cultural communication and traffic monetization by bringing cultural relics to life through AI to better engage young visitors, creating immersive "time travel" photo spots to boost consumption, and providing "turnkey" solutions that reduce IP development cycles by 80%; For consumers, of whom 63% are young users (primarily 18-35 years old), it meets their core needs for immersive "time-travel" in virtual AI-powered try-on experiences, personalized digital creation (clothing, portraits, short videos), and social sharing.

Its core technical solution deeply integrates industry knowledge bases (over 100,000 cultural relic entities, 12,000 patterns and associated graphs), verified by authoritative experts, and overcomes the challenges of cultural misinterpretation in AIGC regarding clothing styles and historical settings. The flexible functional modules, such as dynamic scene generation, AI-powered try-on, pattern extraction, and 4K short video generation, create a closed loop of "experience - photo shooting - consumption." The safe and controllable solution, adapted to Ascend AI processors, has completed dual filings with the Central Cyberspace Affairs Commission and received the highest security certification. While meeting the requirement that cultural relic data be "usable but not visible", it supports private, all-in-one and lightweight SaaS deployment, with full lifecycle support by professional teams.

Its innovative business model incorporates a three-tier system of "hardware access + scenario operation + IP ecosystem". Basic cash flow comes from gradient hardware sales; operational profit-sharing (experience commission, single user fee of 9.9 yuan, short-term rental packages) deepens business binding; IP operations (digital clothing/IP licensing, copyright transaction commission, co-branded cultural creative sales) open up incremental revenue. All this, combined with hardware cost optimization, innovative consumables model and cloud-edge computing power scheduling, achieves 76% gross profit margin.

Experience upgrade
Visitor stay duration at Hangzhou Museum increased by 275%, and the single interactive experience duration at Sichuan Museum rose to 25 minutes.

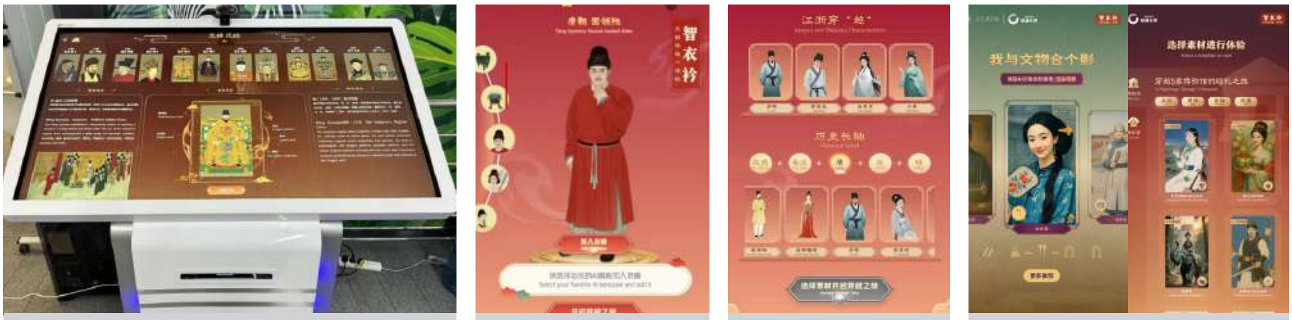
Commercial exposure
The events at the Capital Museum received over 2.3 million views on social media platforms, and Anhui Museum's average daily revenue per device reached 2,850 yuan, with a conversion rate of 35% in derivative products.

Significant results

International popularity
At MWC25 Barcelona, the platform attracted over a thousand experiences in a single day, generating ten thousand digital avatars.

Social value
The platform improves cultural awareness among youth with a penetration rate of 40%. Lightweight SaaS deployment reduces hardware costs for small and medium museums by 60%, and digital costumes reduce carbon emissions to promote green consumption.

With technical precision (large models + expert verification), closed-loop business model (hardware + profit sharing + IP), and social inclusiveness (cultural equality + low carbon), the "Smart Clothing" platform successfully fulfills the vision of "making history tangible, culture accessible, and heritage sustainable", becoming a benchmark for digital innovation in cultural tourism.



v1.0 landscape version

Upload photos and swap faces with materials such as casual wear, imperial robes, and theatrical costumes. Descriptions about costume styles and scenes are available.

v2.0 portrait version

Image + video generation with dynamic material display on homepage and additional AI-powered hanfu feature.

v3.0 portrait version

Added region, venue, and dynasty switching as well as poems composed by DS.

v4.0 museum/exhibition version

Interactive revisions based on actual problems in museum operation, facilitating customer acquisition in museum settings and capability demonstration at exhibitions.

Figure 1 Smart Clothing in diverse versions

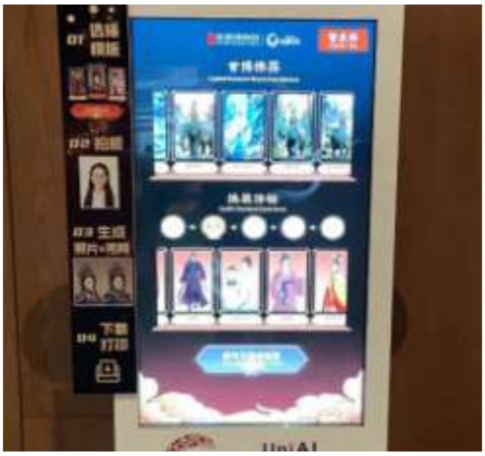
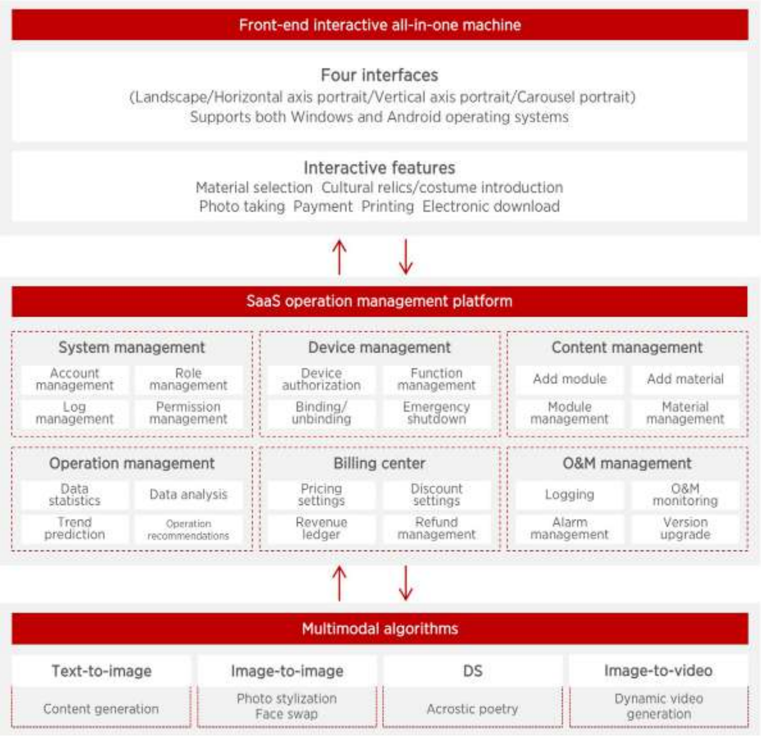


Figure 2 Landscape all-in-one

Figure 3 Portrait all-in-one



Front-end interactive all-in-one machine

Four interfaces
(Landscape/Horizontal axis portrait/Vertical axis portrait/Carousel portrait)
Supports both Windows and Android operating systems

Interactive features
Material selection Cultural relics/costume introduction
Photo taking Payment Printing Electronic download

SaaS operation management platform

System management: Account management, Role management, Log management, Permission management

Device management: Device authorization, Function management, Binding/unbinding, Emergency shutdown

Content management: Add module, Add material, Module management, Material management

Operation management: Data statistics, Data analysis, Trend prediction, Operation recommendations

Billing center: Pricing settings, Discount settings, Revenue ledger, Refund management

O&M management: Logging, O&M monitoring, Alarm management, Version upgrade

Multimodal algorithms

Text-to-image: Content generation

Image-to-image: Photo stylization, Face swap

DS: Acrostic poetry

Image-to-video: Dynamic video generation

Figure 4 Cultural and creative experience modules of Smart Clothing

CHALLENGES

The current cultural tourism industry faces core challenges in digital transformation, which are mainly manifested in four major contradictions: insufficient effectiveness of cultural communication, imbalance between experience and commercial conversion, dormant data assets, and lack of technological adaptability. Specifically, traditional cultural institutions (such as museums and historic districts), while carrying deep cultural value, are constrained by static exhibition modes (90% of county-level museums rely on glass cabinets + text labels), resulting in shallow cultural narratives (72% of armor exhibits are merely labeled as "ancient military attire") and fragmented information. 83% of visitors report "failure to understand the cultural significance of the attire," while 45% of young people misinterpret hanfu shows as merely "photo opportunities at tourist sites," exposing a serious disconnect between professional depth and communication effectiveness. Cultural values are merely exploited for commercial gains. Superficial interactive experiences disappoint 78% of visitors who desire "tactile experiences," resulting in short visitor stays and low conversions of secondary consumption. Museums and other cultural institutions possess massive amounts of precious data assets, but the data often face challenges in terms of privacy, security, domestic production, and being "usable but not visible." AIGC technology has enormous potential, and the market is demanding higher standards for AIGC tools to accurately describe clothing systems, historical scenes, cultural relic patterns and other cultural details, avoiding historical errors or cultural misinterpretations.

The current cultural and creative products are homogenized, with long design cycles and slow market responses, and their IP value can hardly be activated. To break through the dilemmas in cultural and creative product development, the industry urgently needs to solve four major bottlenecks:

- Superficial communication (narrative fragmentation + monotonous forms);
- Fragmented experiences (excessive commercialization + lack of depth in interactions);
- Data silos (dormant digital assets + insufficient activation capabilities);
- Technology misalignment (AIGC content distortion + lack of standards).

Only by integrating technology with culture in a precise and safe manner can we activate cultural values and achieve sustainable monetization.

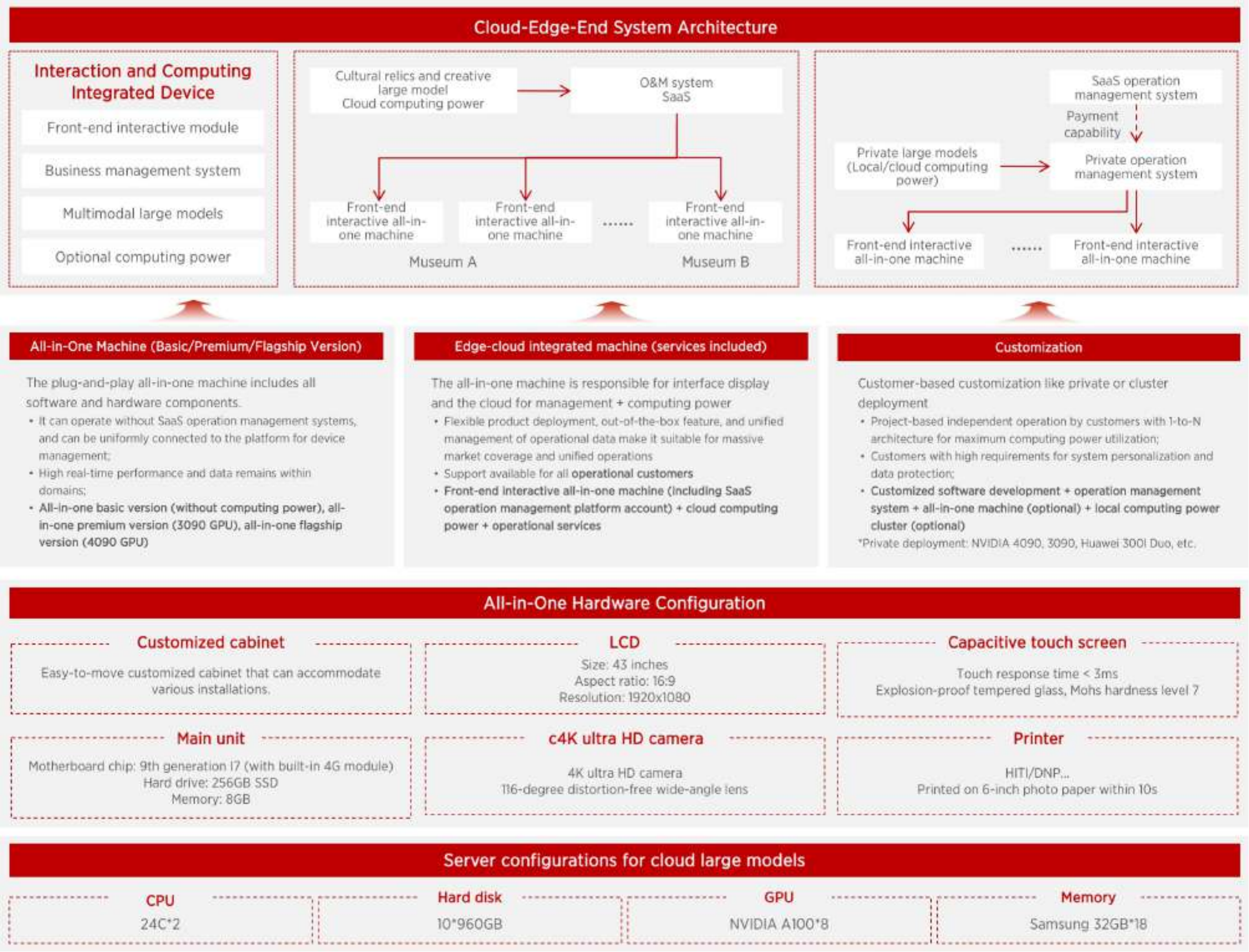
SOLUTION AND VALUE

01 Overall Structure/Solution

Three-tier collaborative architecture

Edge layer (terminal interaction)	AI all-in-one machines (landscape/portrait) integrating HD cameras and GPU computing power (4090 graphics card) are deployed for 4K image generation in 5 seconds and short video output (with special effects/voiceover) in 20 seconds.
Platform layer (intelligent hub)	The UniAI multimodal large model with 204 billion parameters and Chinese-native DiT architecture supports text-to-image/video generation; Chinese CLIP overcomes the 77-token limit and precisely parses long text semantics (such as "Song Dynasty beizi + interlaced lotus pattern"). Security protection system: Dual filings with the Central Cyberspace Affairs Commission and three mechanisms of data quality control, model hallucination suppression, and cultural content review ensure that AIGC output complies with correct values.
Cloud layer (operation and management)	The SaaS platform enables multi-tenant management, material copyright control (pattern authorization), and remote device maintenance, supporting concurrent operation of thousands of devices (peak latency < 0.5 seconds).

With Huawei Ascend AI platform for training and inference, the overall solution supports models with parameters ranging from 7B to 70B.


Cloud-Edge-End System Architecture
Interaction and Computing Integrated Device
Front-end interactive module
Business management system
Multimodal large models
Optional computing power
Cultural relics and creative large model Cloud computing power
O&M system SaaS
Front-end interactive all-in-one machine
Front-end interactive all-in-one machine
......
Front-end interactive all-in-one machine
Museum A
Museum B
SaaS operation management system
Payment capability
Private large models (Local/cloud computing power)
Private operation management system
Front-end interactive all-in-one machine
......
Front-end interactive all-in-one machine

All-in-One Machine (Basic/Premium/Flagship Version)

The plug-and-play all-in-one machine includes all software and hardware components.

- It can operate without SaaS operation management systems, and can be uniformly connected to the platform for device management;
- High real-time performance and data remains within domains;
- **All-in-one basic version (without computing power), all-in-one premium version (3090 GPU), all-in-one flagship version (4090 GPU)**

Edge-cloud integrated machine (services included)

The all-in-one machine is responsible for interface display and the cloud for management + computing power

- Flexible product deployment, out-of-the-box feature, and unified management of operational data make it suitable for massive market coverage and unified operations
- Support available for all **operational customers**
- **Front-end interactive all-in-one machine (including SaaS operation management platform account) + cloud computing power + operational services**

Customization

Customer-based customization like private or cluster deployment

- Project-based independent operation by customers with 1-to-N architecture for maximum computing power utilization;
- Customers with high requirements for system personalization and data protection;
- **Customized software development + operation management system + all-in-one machine (optional) + local computing power cluster (optional)**

*Private deployment: NVIDIA 4090, 3090, Huawei 300I Duo, etc.

All-in-One Hardware Configuration

Customized cabinet
Easy-to-move customized cabinet that can accommodate various installations.

LCD
Size: 43 inches
Aspect ratio: 16:9
Resolution: 1920x1080

Capacitive touch screen
Touch response time < 3ms
Explosion-proof tempered glass, Mohs hardness level 7

Main unit
Motherboard chip: 9th generation I7 (with built-in 4G module)
Hard drive: 256GB SSD
Memory: 8GB

c4K ultra HD camera
4K ultra HD camera
116-degree distortion-free wide-angle lens

Printer
HITI/DNP...
Printed on 6-inch photo paper within 10s

Server configurations for cloud large models

CPU	Hard disk	GPU	Memory
24C*2	10*960GB	NVIDIA A100*8	Samsung 32GB*18

Figure 5 China Unicom UniAI "Smart Clothing" cloud-edge-end architecture

02 Application Scenarios

Scenarios	Solutions	Cases
Museum special exhibitions	3-day rapid customization of digital cultural relics (such as Lady Fu Hao's owl zun -> bronze pattern attire)	Ten sets of custom costumes for the Shang Dynasty Exhibition at the Capital Museum, driving visitor interaction rate up by 40% ↑
Cultural scenic areas	Take photos to create "time travel" short videos (Leifeng Pagoda of Hangzhou West Lake + Lady Bai Suzhen image)	Special exhibition at Hangzhou Museum with conversions of secondary consumption increased by 50% ↑
International exhibitions	Lightweight SaaS deployment and multilingual generation of Chinese-style portraits	MWC Barcelona 2025 with 1000 visitor experiences in a single day
Cultural and creative development	Millisecond-level pattern extraction to derivative product design	Cultural relic pattern reuse rate increased from 12% to 85%

03 Key Innovations

Cultural precision engine

- Cultural relic database with million entries + expert rule database: 10,000+ cultural relic entities, 12,000 traditional patterns, and 3,000+ form and system rules calibrated by National Museum experts (for example, "Tang Dynasty veiled hats are prohibited in Song Dynasty costumes").
- Multimodal association technology: user photos -> smart matching with cultural relic patterns -> generation of short videos with historical narratives (example: uploading a photo and entering "Yue Fei" to generate dynamic scenes of "Marshaling Troops on the Battlefield").

Closed-loop production and operation

- IP activation link: cultural relics -> digital costumes -> short videos -> one-click derivative creation (for example, the design of refrigerator magnets based on Sichuan Museum pattern is shortened from 2 weeks to 2 hours).
- Data-driven operations: Real-time analysis of user preferences (such as popular dynasty costumes), producing "visitor portrait reports" to guide IP matrix construction.

Flexible cloud-edge-end deployment

- Private all-in-one machine: meeting data security requirements (such as local processing of museum data).
- Lightweight SaaS: out-of-the-box for quick launch in scenic spots (0.5-hour deployment).

04 Business Models

Model	Charging method	Target customers	Cases
Hardware sales	All-in-one machine priced at 150,000-300,000 yuan/unit	Large museums/4A scenic spots	Customized device for Sichuan Museum
SaaS subscription	Annual fee of 80,000-200,000 yuan (based on number of terminals)	Small and medium cultural venues/ temporary exhibitions	Rental for special exhibition at Hangzhou Museum
Value-added services	Custom material fee: 5000 yuan/set	Government cultural projects/ commercial cultural tourism IP	10 sets of custom costumes for the Shang Dynasty Exhibition at the Capital Museum
Profit-sharing cooperation	15%-30% revenue share from derivative products sales	Cultural and creative developers	Co-branded products with Anhui Museum achieve sales of 100,000 units

05 Core Values (including quantitative results and qualitative impact)

Quantitative results	
Experience enhancement	Visitor stay prolonged by **30%** (from 15 minutes to 20 minutes), interaction rate improved from < 5% to **35%** (Capital Museum Shang Dynasty Exhibition).
Operational efficiency improvement	Creative design cycle shortened by **90%** (after reduction, 10 sets of costumes can be customized in 3 days), cultural relics data reuse rate increased from **12%** to **85%**.
Business growth	Conversions of secondary consumption increased by **50%** ↑ (Hangzhou Museum special exhibition), and derivative product development cost reduced by **70%**.

Qualitative impact	
Avoiding cultural misinterpretations	AIGC error rate **< 3%** (compared to industry average of 40%), correcting cognitive biases such as "fengguan xiapei attire is traditional wedding dress".
Activating data assets	Transforming dormant artifacts into wearable digital IPs (e.g., Lady Fu Hao's owl zun pattern -> short videos on social platforms).
Supply chain sustainability benchmark	full-stack adaptation to the Ascend platform.

REFLECTION

Key Insights

Technology-driven cultural revitalization Precision breakthrough

- The self-developed multimodal large model UniAI (204 billion parameters) and 100,000+ cultural relic database address cultural misinterpretation in AIGC (costume style error rate < 3%), ensuring authoritative historical restoration.
- Immersive experience innovation: Innovative interaction with relics through time and space prolongs visitor stay by 275% (Hangzhou Museum) through such features as AI costume try-on and dynamic scene generation (5 seconds per image/30 seconds per video).

Sustainable business model

- Three-tier profit model: Hardware sales (affordable version 150,000 yuan/flagship version 200,000 yuan) + operational revenue sharing (tickets/single experience 9.9 yuan) + IP ecosystem (digital costume licensing, co-branded cultural creative profit sharing), with comprehensive gross profit margin exceeding 76%.
- Lightweight deployment: Private all-in-one machine ensures data security, and SaaS model lowers the digital threshold for small and medium museums (cost reduction by 60%).

Social value

- Cultural awareness among youth: AI costume try-on with cultural relic patterns and interactive historical scenes better engages young visitors. The penetration rate increased by 40%.
- Green low-carbon practice: Digital costumes replace physical hanfu rentals, reducing carbon emissions from the textile industry (daily revenue of 2850 yuan per machine at Anhui Museum).

Future Plans

Technology development

- 2025: Develop "Cultural Relic Holographic Projection" module, integrating AR technology to achieve multi-sensory interaction.
- Build an open API platform to support third-party developers in customizing cultural IP (such as digital blind boxes, virtual idols).

Massive replication

- Business penetration: Cover 50+ top-tier museums within 3 years (currently in partnership with 10 museums including the Capital Museum, Sichuan Museum, etc.), and reach county-level cultural venues (with 80% policy coverage).
- Consumer viral growth: Launch social sharing incentive system, where user-generated content can be exchanged for co-branded cultural products (target: 60% of derivative product revenue).

Ecosystem co-building

- Establish "Museum AIGC Application Standards" with the National Cultural Heritage Administration to deliver industry paradigms of "technology + rules."
- Develop the "Belt and Road" digital cultural tourism scenarios (verify international acceptance at MWC) to promote Chinese culture globally.

GSMA

1 Angel Lane
London
EC4R 3AB
United Kingdom

Tel:+44 (0)20 7356 0600
Fax:+44 (0)20 7356 0601