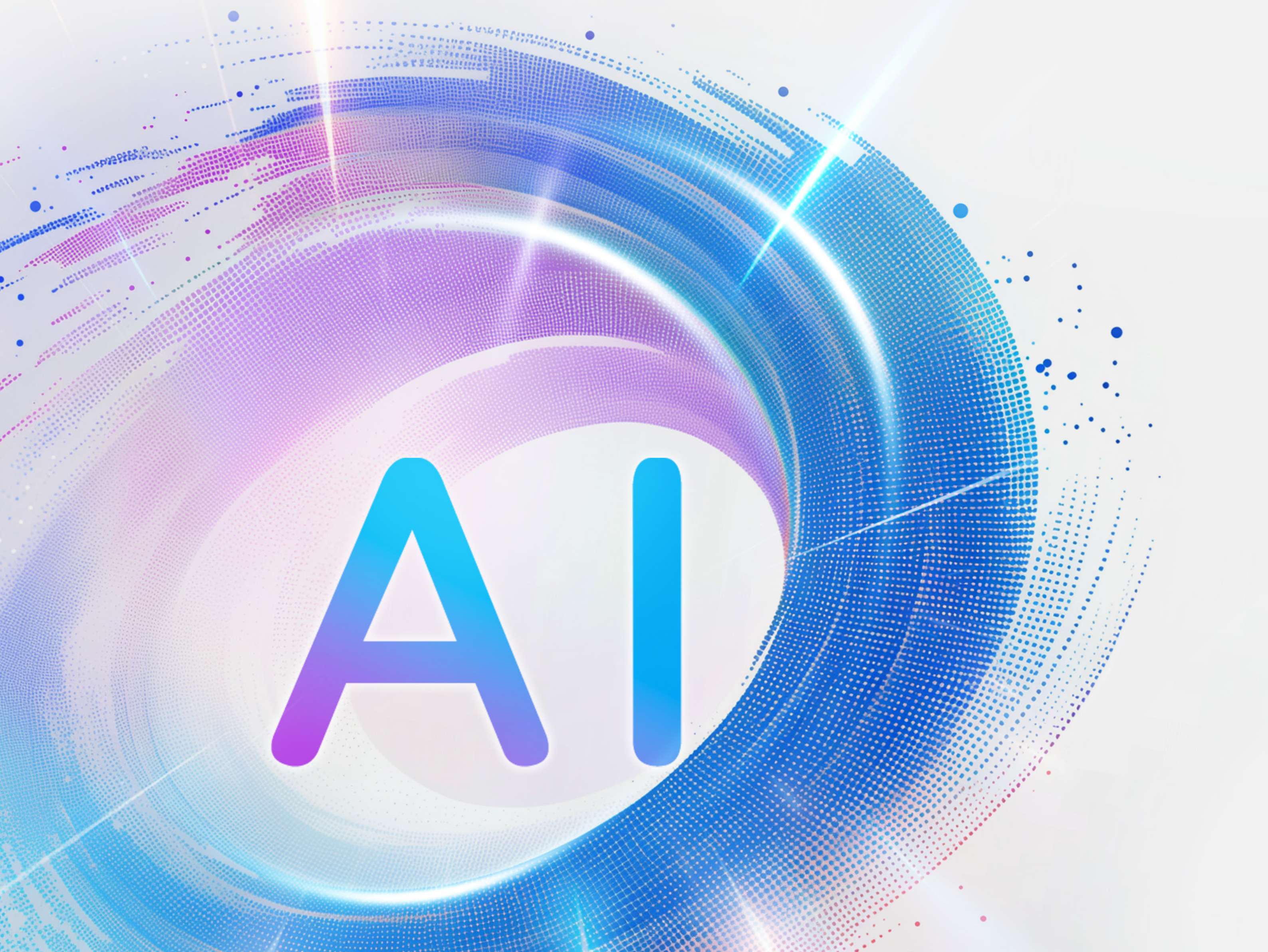
ZTE

2025
ZTE Autonomous Networks White Paper

Value-Driven, AI-Innovation: Opening a New Chapter for Highly Autonomous Networks



Contents

Abstract

Cornerstone and key enablers of the digital world, telecom operators are accelerating the construction of highly automated, intelligent networks integrating cloud, network, computing, and Artificial Intelligence (AI). Autonomous networks, characterized by automation and intelligence, are poised to become an endogenous capability of future communication systems across all industries.

As a proactive advocate and provider of autonomous networks, ZTE actively contributes to the industry transformation - from standardization and capability building to innovative practices. This White Paper outlines ZTE's strategy to support operators in their autonomous networks journey. It elaborates on ZTE's core philosophy of being driven by value outcomes, leveraging openness and decoupling alongside digital intelligence engines to accelerate the resolution of bottlenecks and disruptions, ultimately achieving end-to-end automated and intelligent O&M. It also highlights ZTE's exemplary achievements in advancing towards Level 4 highly autonomous networks.

This White Paper presents and discusses :

Insights into the intelligent network era

Summarizes the main aspects of network evolution, providing perspectives on future business and autonomous networks.

ZTE's AIR Net highly autonomous networks solution

Describes ZTE's approach to breaking bottlenecks in end-to-end automation using data and capability decoupling, leveraging key technologies like large and small model collaboration, agents, and digital twins. It addresses automation challenges, reduces the skills threshold for network operations, and gradually replaces manual intervention with autonomous closed-loop networks. Grounded in value scenarios and tangible outcomes, it ensures a closed loop of commercial success. The solution employs a modular digital architecture, powered by a digital intelligence engine that delivers diverse service capabilities. Coupled with scenario-specific autonomous intelligence, it enables rapid orchestration and deployment while supporting future application paradigm evolution.

Development trends of autonomous networks

Offers in-depth insights into the development trends of the autonomous network industry and its challenges, based on the evolution of current standards and industrial practices. It explores advancements in areas such as digital human live streaming, 5G-A business development, the integration of cloud, network, computing, and intelligence, as well as new technologies and applications like LLMs, agents, and digital twins. Furthermore, it offers policies and recommendations to address these challenges.

Success stories

Shares highly autonomous network case studies in automation, intelligence, and energy efficiency optimization, offering valuable references for global operators and partners.

Industry collaboration proposal

Calls for industry partners to work together in a spirit of co-creation and mutual benefit to advance autonomous networks to higher levels.

01

Insights into Intelligent Network Era

This section introduces the main ongoing developments for Autonomous Networks in the industry, and the overall implications for telecom operators.

1.1 Autonomous Network Standards and Industry Practices Overview

The Development of Autonomous Network Standards

Several international standardization organizations, such as TM Forum, NGMN, ITU, 3GPP, and ETSI, are actively advancing autonomous network standards, including architectures, autonomy levels, value scenarios, outcomes, and intelligent technologies. These organizations integrate the needs of operators and the industry to develop the autonomous network ecosystem. As a key milestone, 3GPP released the first autonomous network level standard, TS 28.100, in its R17 version, providing a critical evaluation framework for assessing the levels of network automation and intelligence within the telecommunications industry.

The China Communications Standards Association (CCSA) has also made significant progress in autonomous network standardization, spearheading the development of related standards and research initiatives.

Looking forward to Level 4 and Level 5 autonomous networks, TM Forum envisions enabling technologies such as network AI LLMs, network trustworthiness, and digital twins. These technologies will support the endogenous data, algorithms, and computing power requirements for AI in 3GPP networks, driving the evolution of autonomous networks toward higher levels of intelligence.

ZTE has played a leading role in the development of autonomous network standards, staying at the forefront of technology. As illustrated below, ZTE actively promotes autonomous network standards within organizations like TM Forum, 3GPP, ETSI, and CCSA, contributing to the industry's progression toward high autonomous networks.

Figure 1. ZTE's Participation in Autonomous Network Standard Developing Organizations

Industry Practices in Autonomous Networks

Chinese operators have demonstrated remarkable leadership in both domestic and international autonomous network practices. The three major operators in China have incorporated autonomous networks into their respective Group strategies, setting goals to reach Level 4 autonomous networks by 2025 and achieving substantial progress in application innovation.

China Mobile: China Mobile has established a comprehensive theoretical framework for Autonomous Networks and conducted in-depth explorations in autonomous practices, positioning itself as a benchmark for global operators. By focusing on high-value scenarios, it has integrated previously isolated automation capabilities into end-to-end fully automated processes, creating streamlined intelligent O&M and realizing significant "quality improvement, cost reduction, and efficiency enhancement."

China Telecom: China Telecom has implemented its cloud-network convergence strategy, introducing the "Three-Pole Vision"based on:Ultimate Customer Experience, “Self-X”Ultrafast Product Services (Self-Sensing, Self-Loading, Self-Marketing, Self-Activation, Self-Planning, Self-Healing, Self-Maintenance, Self-Optimization, Self-Accounting), “Zero-X”Supremely Intelligent Cloud-Network Operations (Zero Touch, Zero Wait, Zero Faults), objectives and publishing a roadmap for autonomous cloud-network operations to drive digital transformation.

China Unicom: China Unicom has proposed a "Three Transformations (Digitalization, Intelligence, Agility), Three Layers (Network Layer, Service Layer, Business Layer), Three Closed Loops (Knowledge Closed Loop, Task Closed Loop, Intent Closed Loop)" architecture, combined with the "Four-Zero-Four-Autonomy" vision and precision commercial value initiatives, advancing network intelligence and operational transformation.

Internationally, TM Forum, in collaboration with numerous industry partners, released a white paper and produced the Autonomous Networks Manifesto, signed by ZTE in [...], marking a new milestone in the global development of the autonomous network industry.

Vodafone: Focuses on Level 4 pre-emptive automation, leveraging AI for automated and predictive maintenance and optimization in fiber and cable scenarios.

Deutsche Telekom: Targeting Level 4, Deutsche Telekom introduced the concept of Dark NOC, envisioning a highly autonomous network operation center that functions without human intervention. The goal is to achieve a fully unmanned, digitalized Network Operations Center.

MTN (South Africa): Focuses on fault management and IP quality optimization scenarios, advancing innovative practices in autonomous networks while simultaneously implementing process transformation and capability enhancement.

AIS (Thailand): Incorporates autonomous networks into its "Cognitive Techco" strategy, aiming for Level 4. It emphasizes high-value scenarios such as customer complaint management, network optimization, and configuration changes. By leveraging advanced data analytics, high automation, and AI technologies, AIS accelerates its digital and intelligent transformation and improves its AN maturity levels.

Telkomsel (Indonesia): Positions autonomous networks as a core element of its digital innovation strategy, with a goal to achieve Level 4 by 2025. The focus is on maintaining service quality during peak periods.

Global operators are accelerating their transition to network automation and intelligence, striving to reach high autonomous networks. This transition aims to enhance network efficiency, improve user experience, and enable network service monetization.

1.2 Operators Transformation Journey in the Network Intelligence Era

AI technology accelerates the development of digital and intelligent transformation in society. This transformation is profoundly changing people's modes of production and life while reshaping the global economic structure and social operation models.

Characteristics of the Era of Intelligence

New information infrastructure

The new information infrastructure provides fundamental digital services such as sensing, transmission, storage, and computing for the digital economy. It becomes the foundational capability and innovation driving force in the era of intelligence.

Business innovation

The advent of LLMs had a disruptive impact on business, significantly altering interaction paradigms. Multi-modal content can now be generated in real time, unlocking a variety of creative application scenarios and providing more immersive business experiences and personalized services.

Bridging the intelligence gap

The era of intelligence will enable everyone to create AI systems, gaining access to powerful AI models and computing resources without the need for advanced mathematics and computer science skills. This will mean that more people can use AI for research and product development in their respective fields, greatly lowering the technological barriers to AI.

Reshaping production methods

The application of LLM technologies will free people from tedious tasks, allowing them to focus on work goals and business value, greatly enhancing productivity. Data has already become a key production factor in the digital economy era, and new quality productive forces will be the key driving force behind the development of the era of intelligence. The social division of labor will also undergo fundamental changes.

Transformation of Operator Networks in the Era of Intelligence

The transformation in the era of intelligence is a complex process, marked by both opportunities and challenges. Telecom operators need to undergo profound changes to adapt to new development trends.

Operators need to provide ubiquitous and reliable connectivity to achieve an intelligent, integrated digital information infrastructure that embodies "high-speed ubiquity, seamless integration of space and ground, and the integration of cloud, network, computing, and intelligence," in order to meet the high-quality network requirements of the unmanned economy.

In response to new consumption patterns, telecom operators need to provide timely and personalized experiences. By establishing a comprehensive data management and security assurance system throughout the entire lifecycle, telecom operators can create traffic entry points and multi-touch interconnectivity that are technological, intelligent, and ecological, meeting the diverse needs of users.

For the intelligent development of social governance and services, telecom operators need to provide accessible and convenient services while protecting privacy and preventing ethical risks.

In this transformation process, operators rely on their network infrastructure to address the root problems of intelligence by providing data services. They will use LLMs to free up human labor and simulate human organizational behavior. Additionally, digital twin technology will enable the mapping, simulation, and closed-loop control of physical environments to virtual environments. These three elements together form the comprehensive AI transformation necessary for achieving autonomous networks, thus driving networks toward highly autonomous network evolution.

From an Operations and Maintenance (O&M) perspective, these will involve the following six dimensions.

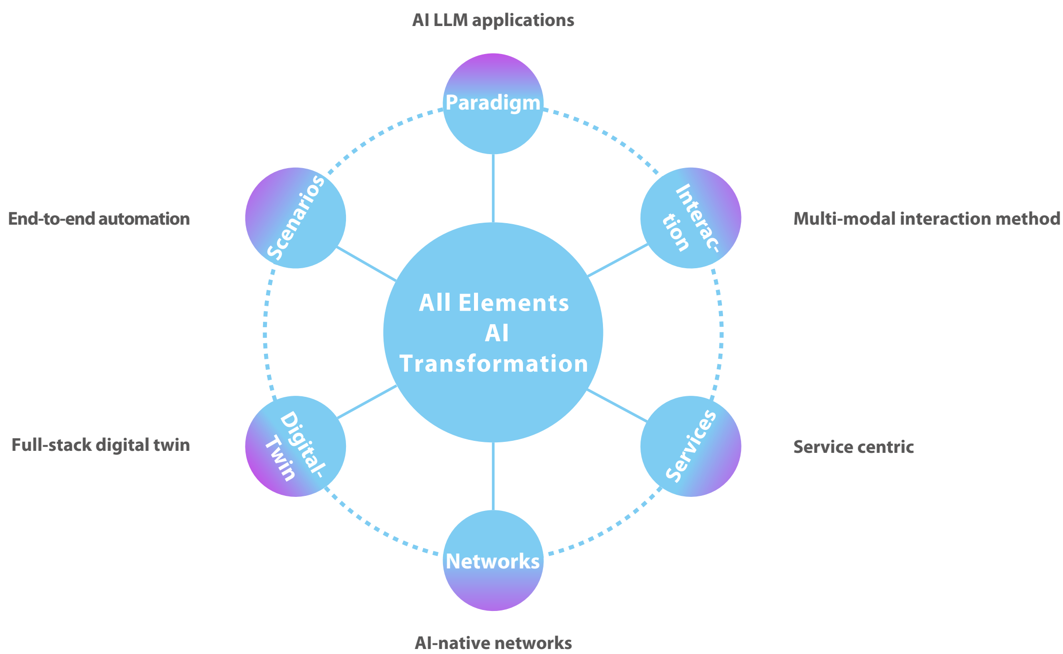


Figure 2. AI Components for Autonomous Networks Evolution

Services

The transformation of operations and maintenance (O&M) from a network-centric to a user experience and business experience-centric approach will also make network services more personalized. The network will be able to automatically adjust based on user behavior and preferences to provide the best experience.

Scenarios

Successful transformation dictates to focus on high value scenarios exhibiting convenient Return on Investment (ROI), and then continuously improving end-to-end automation and intelligence capabilities. Through data openness, capability openness, and the application of LLMs and agents, the automation of O&M and operations can be realized stepwise in various scenarios, addressing key processes such as business provisioning, network changes, fault processing, network planning and optimization, etc.

Interaction

By integrating LLM technologies, the network can understand and respond to various inputs, such as voice and video, providing users with a richer, simpler and more natural interaction experience.

Network

AI-enabled network elements and applications further enhance endogenous intelligence capabilities. By automatically adjusting configurations based on real-time data and predefined policies, the self-learning, and self-optimizing capabilities are enhanced, together with network performance, service quality and customer experience.

Digital Twins

digital twin creates a digital replica of the network, enabling real-time monitoring, fault prediction, and performance optimization of the network status, as well as customer experience modeling and prediction.

Paradigm

The introduction of Agentic AI allows to create multiple agents (such as for monitoring, fault correction, optimization, etc.) which collaborate automatically and intelligently within the network to solve complex problems.

These six aspects continue to evolve, allowing operators to provide more efficient and intelligent services, leading to fully intelligent network operations, and creating more value for society and industries.

02

Development Trends of Autonomous Networks

Network operators are facing challenges from new business opportunities, while new technologies are also providing transformative means for network evolution. Autonomous networks are gradually revealing new trends in areas such as new businesses, network evolution, emerging technologies, and applications.

2.1 Trend 1: Rapid Growth of New Businesses Bring New Monetization Opportunities

Insight and Analysis

“Digital humans”create connections on an emotional level with the look, voice, and personality of a real human, helping businesses to build trust, improve the customer experience, and promote brand awareness. Digital human live streaming business uses AI technology to create virtual hosts or characters with autonomous speaking and behavioral capacities. This form of live streaming no longer depends on actual hosts but instead employs digital technology to enable virtual performances. According to the White Paper on the Development of China's Virtual Digital Human Industry in 2024 published by iiMedia Research on April 19, 2024, China's virtual human-driven industry market size and core market size reached RMB 333.47 billion and RMB 20.52 billion respectively in 2023. These figures are projected to grow to RMB 640.27 billion and RMB 48.06 billion by 2025, demonstrating robust growth momentum.

The 2024 JD.com "6·18" report shows that the number of merchants launching digital human live-streaming rooms increased by nearly 400% compared to last year's "Double 11 Shopping Festival". The advantage of AI digital humans lies in their ability to live stream for extended periods, providing continuous information broadcasts, which can significantly increase brand exposure while reducing labor costs and operational expenses.

Metaverse business is also showing strong development momentum. According to the 2024-2029 China Metaverse Industry Development Deep Dive and Future Trends Forecast Report released by the CIRN, the Metaverse market is continuously expanding, with the domestic market expected to surpass RMB 280 billion by 2025. 5G technology large-scale coverage, and virtual reality technologies such as AR and VR meet the foundational requirements for the Metaverse, and blockchain technology is steadily advancing to provide technical support for digital asset authentication and transactions in the Metaverse. According to iiMedia Consulting data, 62.4% of internet users are willing to engage in Metaverse social interactions, reflecting a high level of interest and willingness to participate in the Metaverse.

Telecom operators face numerous challenges when supporting these emerging businesses. Network stability is fundamental, while the clarity and quality of real-time video transmission directly impact user experience. Regarding the Metaverse, the Metaverse White Paper (2023) released by the China Academy of Information and Communications Technology (CAICT) states that the Metaverse transmission network must focus on low latency, high stability, and high quality for global cross-domain deterministic communication capabilities.

Recommendations

To address the challenges posed by AI digital human live-streaming and Metaverse business, telecom operators can adopt the following policies :

Strengthen network infrastructure construction

To cope with the rapid growth in traffic, operators should foster 5G business and terminal innovation, and promote ecosystems and collaborations in areas like ultra-high-definition, cloud VR, cloud AR, and cloud gaming.

Build computing power networks

Relying on cloudification and digital transformation strategies, operators can construct new cloud-network-intelligent computing infrastructure that supports the Metaverse, offering immersive services based on XR and AI technologies.

Automated sensing and analysis capabilities

Using AI and big data analytics, operators can achieve real-time monitoring and intelligent optimization of cloud, network, and computing infrastructures. For example, network slicing technology can dynamically schedule traffic based on different scenarios to ensure optimal resource allocation. Additionally, combining AI monitoring tools to analyze real-time network traffic evolution allows for automatic detection of network bottlenecks and optimization adjustments, ensuring user SLA compliance and further improving user experience.

Design new business models

To enhance operator revenue and accelerate commercial monetization, operators should tailor solutions for high-value customers and specific business scenarios.

2.2 Trend 2: 5G-A Accelerates Communication and Sensing Integration, Promoting Low-Altitude Economy Development, and Bringing New O&M Challenge

Insight and Analysis

The low-altitude economy (which refers to economic activities in airspace below 1000 meters) has become a strategic emerging industry (source: Central Economic Work Conference – dec. 2023). The 2024 Chinese Government Work Report emphasized the active development of new growth engines, including biomanufacturing, commercial aerospace, and the low-altitude economy. On November 18, 2024, China’s Central Air Traffic Management Commission announced the launch of eVTOL (electric vertical take-off and landing) pilot projects in six cities. The initial six pilot cities have been tentatively identified as Hefei, Hangzhou, Shenzhen, Suzhou, Chengdu, and Chongqing. The pilot documents outline plans for flight routes and areas, as well as delegating airspace below 600 meters to local governments. The second batch of pilot cities will be announced soon, and the utilization rate of domestic low-altitude airspace is expected to rise, signaling the industry's entry into a rapid growth phase.

5G-A communication and sensing integration support the rapid development of the low-altitude economy, while also opening up vast commercial potential for telecom operators. Firstly, operators can build low-altitude communication and sensing networks to provide infrastructure for the low-altitude economy. Secondly, telecom operators can drive technological innovation in this field, expanding application scenarios and participating in the operation of low-altitude flight service centers and the formulation of low-altitude traffic regulations. Additionally, by establishing low-altitude economy industry alliances, operators can coordinate efforts from industry, academia, and research institutions to jointly foster the industry ecosystem. At the same time, operators should gain a deep understanding of the low-altitude economic demands of key industries and offer customized solutions to enrich the application scenarios within the low-altitude economy.

Recommendations

The development of the low-altitude economy presents new business opportunities and challenges for telecom operators:

- **High-precision detection and efficient data processing:** The new business demands from the low-altitude economy are driving the construction of integrated communication and sensing networks. These networks not only provide communication services but also possess sensing capabilities to achieve 3D coverage and sensing. Basic networks, such as base station upgrades, must integrate communication and sensing service capabilities to meet the needs of high-speed drone flight and inter-base station operations. The construction of sensing-integrated networks involves technologies such as large-angle, continuous waves, pulse waves, multi-cell joint de-duplication, and millimeter waves, which help achieve high-precision detection and sensing. Moreover, a network management system is needed to efficiently and quickly process drone flight data and deliver comprehensive sensing services.
- **Construction of low-altitude communication and sensing security management systems:** Security is the bottom line for the development of the low-altitude economy. Telecom operators must find core application scenarios for the low-altitude economy based on visibility, control, and traceability. To address security issues related to low-altitude aircraft, operators need to enhance their autonomous capabilities in critical technologies to ensure industrial and information security. At the same time, operators must find a balance between "release" and "control"—ensuring the security of low-altitude regions while unleashing the potential of the low-altitude economy. This involves user information registration, the construction of flight permit issuance platforms, and the improvement of low-altitude airspace management systems.
- **Exploration of low-altitude operation models:** Telecom operators need to develop full-lifecycle operational caperformance. In particular, the drone logistics and inspection market, with its vast market potential, is expected to bring significant revenue growth for operators.

As wireless networks transition from traditional ground coverage to three-dimensional coverage that spans both the ground and low-altitude regions, low-altitude O&M capabilities face new challenges. Due to the small size and low altitude of low-altitude targets, combined with the complex electromagnetic environment, these factors place higher demands on network planning and maintenance, prompting a need for precise simulations and real-time, intelligent network optimization technologies.

Abilities to support both connection services and sensing innovation services. In fields such as aviation emergencies, traffic inspections, and logistics transportation, the application scenarios of the low-altitude economy are continuously expanding, providing new growth points for operators' .

2.3 Trend 3: Growth in Intelligent Computing Scale, Integration of Computing and Network, and Ongoing Industry Efforts to Enhance Resource Utilization

Insight and Analysis

Rapid growth in intelligent computing scale : Driven by the Scaling Law, the scale and capabilities of LLMs are expanding rapidly, with the demand for training computing power accelerating significantly. The size of individual training clusters is progressing from tens of thousands of GPUs to hundreds of thousands. However, the pace of AI chip advancements is lagging behind the growth in cluster computing power demands, further intensifying the expansion of training cluster sizes. Large-scale training clusters introduce immense computing, data communication, and storage pressures, posing severe challenges to cluster stability. Hardware faults, node performance degradation, and network fluctuations can lead to frequent training interruptions, severely impacting the training process, reducing resource utilization, and causing energy waste. For example, Meta's report on the training of its OPT-175B model indicated a total of 90 restarts during the training process.

AI innovation applications as investment hotspots : With the advancement of LLM capabilities, agent-centric AI applications are emerging as a new investment focus, continually creating new AI scenarios and increasing societal acceptance of AI. For instance, OpenAI reported 300 million weekly active users by the end of 2024. These trends are driving the rapid growth of AI inference demand. Simultaneously, the proliferation of agent-based applications inevitably leads to a surge in network connections, signifying a transition from the eras of human-to-human and IoT connections to the era of agent connectivity. These developments present greater challenges for inference costs and the integration of computing and network.

Recommendations

Autonomous intelligent computing networks: The development of autonomous networks to meet intelligent computing training requirements involves two key aspects: first, leveraging ultra-large-scale GPU cluster computing power as the foundational infrastructure for cloud resource pools; second, enhancing the intelligent O&M capabilities of GPU clusters. Pre-training requires rapid health stress testing, including pre-stress testing, inspections, pre-checks, and fault prediction services. During training, the system must have the capability to monitor cluster health in real-time, detecting issues such as GPU card faults, RDMA network faults, node faults, and Pod status anomalies. The system should also support rapid fault detection and real-time visualization. In the event of a fault, it must achieve fault demarcation and positioning within minutes, enable job rescheduling, and handle node isolation, replacement, or self-healing to resume training seamlessly. Additionally, checkpoint read/write performance must be optimized to quickly complete interrupted training. These measures, through end-to-end automated and intelligent O&M, enhance resource utilization and reduce O&M costs.

Inference, alongside training, is another critical capability for autonomous networks. By leveraging breakthroughs in technologies such as knowledge distillation, model quantization, and inference engine optimization, inference costs can be continuously reduced. Models of different sizes can be selected based on application-specific requirements to balance inference hardware, speed, and performance. These measures enable low-cost inference, supporting the development needs of autonomous networks.

2.4 Trend 4: LLMs Empower Operators' Digital and Intelligent Transformation, Accelerating Highly Autonomous Networks

Insight and Analysis

Rapid development of GenAI: In the communication industry, the emergence of new business scenarios such as the low-altitude economy, IoT, connected vehicles, and immersive XR presents telecom operators with multiple challenges. These include effectively controlling operational costs, meeting the continuously rising demands for user experience, and addressing increasingly complex network architectures. At the same time, the ongoing technological revolution, particularly the rapid advancement of GenAI, is becoming a key driver of new quality productive forces for enterprises. According to Gartner, the proportion of enterprises applying GenAI-related technologies in production environments will surge from 5% in 2023 to over 80% by 2026. Additionally, technologies like LLM, prompt engineering, and multi-modal capabilities are expected to mature within the next 2 to 5 years, further accelerating the pace of industrial intelligence.

In this context, the following core technologies are fundamentally transforming the ways telecom operators manage and operate networks:

Agents

Agents have emerged as one of the core technologies of GenAI in 2024. With autonomous learning, task execution, and multi-task collaboration capabilities, agents provide telecom operators with a new pathway for intelligent transformation. By deploying agents, operators can achieve highly automated network management and optimization, enhancing self-sensing and self-repair capabilities within networks. This leads to significantly reduced O&M costs and improved network reliability.

Multimodal LLMs

Multimodal LLMs, capable of processing text, images, and audio, create more immersive interaction methods for customer service. By integrating multi-modal technologies, operators can deliver personalized solutions in customer service, optimize user experiences, and enhance service value.

Inference enhancement

In September 2024, OpenAI introduced the O1 model, which demonstrated exceptional performance in complex inference tasks, establishing a new paradigm of "slow thinking" for LLMs. This model and its enhanced inference capabilities provide novel solutions for communication network O&M and operation. For example, the O1 model and its enhanced inference technologies can enable task orchestration, problem analysis, flexible decision-making, and optimized execution in complex business scenarios. These technologies automatically identify root causes and generate solutions, effectively reducing manual intervention and increasing the automation rate of O&M.

Large and small model collaboration

Traditional AI algorithms and small models continue to excel in areas like data analysis and prediction, offering advantages in accuracy and cost-effectiveness in the short term. By combining GenAI with traditional AI through large and small model collaboration, operators can effectively accelerate the path to achieving Level 4 autonomy.

Recommendations

As AI/GenAI technologies continue to advance, autonomous networks are progressing toward the goal of Level 4 full-stack AI embedment. By embedding AI/GenAI technologies across network element layers, management layers, service layers, and business layers to achieve automation and intelligence, telecom operators are advised to establish the following key capabilities:

- **Selection of network O&M scenarios:** Through human-machine collaboration, large-small model collaboration, and multi-agent collaboration, it is recommended to focus on scenarios such as network fault processing, complaint resolution, change monitoring, and business quality optimization to significantly enhance O&M efficiency and reduce operational expenditure (OPEX).
- **Service quality and operational experience:** Strengthen the security and reliability of services, optimize user experience, and improve customer satisfaction.
- **Business innovation and value expansion:** Leverage AI-driven insights and optimization capabilities to explore new business domains, create greater customer value, and help operators maintain a competitive edge.

In summary, the rapid evolution of GenAI technologies not only aids telecom operators in addressing complex challenges but also propels them from automation to intelligence, enabling the leap to the Level 4 high autonomous stage.

2.5 Trend 5: Digital Twins Accelerate the Evolution of Network Intelligence, Realizing Virtual Representation and Control of the Real World

Insight and Analysis

Digital twin technology has been widely applied in manufacturing, logistics, energy, and other fields. By creating virtual models of physical assets, it enables real-time monitoring, predictive maintenance, and optimized management, improving production efficiency, reducing costs, and enhancing product quality. Particularly in the context of intelligent manufacturing and Industry 4.0, digital twin technology has emerged with new features such as real-time data integration, lifecycle management, adaptive optimization, intelligent decision support, and high-level collaboration.

In the communication industry, the application of digital twins is mainly focused on network construction, optimization, and operations. By building virtual network models, operators can simulate real-world network environments to perform fault prevention, performance optimization, and resource scheduling, enhancing network adaptability.

However, the development and evolution of digital twin technology in the communication industry also face several challenges, such as :

- How to quickly process and respond to changes in high-frequency, large-scale data streams to ensure that the digital twin model accurately reflects network status and makes timely optimization decisions;
- How to ensure that the digital twin model accurately simulates and reflects the real network environment, including dynamic network payloads, equipment statuses, and user behaviors, to maintain high precision and reliability;
- How to effectively integrate data from various sources, formats, and frequencies and extract valuable information from heterogeneous data in large-scale network environments.

As networks progress toward more high autonomous stages, operators' expectations for the automation and intelligence of network operations are increasing, with a stronger focus on realizing commercial value. In the future, digital twins will continue to be a key technology concept for achieving high autonomous networks.

Recommendations

For the various components, including sensing, analysis, decision-making, and execution, research should focus on empowering accurate evaluation and decision-making, enhancing the level of network intelligence, and providing comprehensive support across network planning, construction, maintenance, optimization, and operation. By achieving collaboration across different stages, and continuously optimizing sensing technologies and data analysis models, operators can construct digital twin networks to achieve low-cost virtual-real mapping and control. Coupled with intelligent decision systems, this approach ultimately ensures adaptive and self-optimizing capabilities for networks, ensuring efficient, flexible, and precise network management and operations.

Multi-objective collaborative optimization: By setting multiple optimization goals and coordinating the relationships between various objectives and metrics, this approach ensures that local optimization is achieved without compromising overall global performance, thereby enhancing the overall system efficiency. Using intelligent algorithms and refined modeling, the optimization balance can be achieved across different goals (such as network capacity, latency, and energy efficiency), avoiding local optimizations that may degrade overall performance. This effectively addresses issues such as resource allocation and load balancing in complex networks, ensuring that all subsystems achieve optimal performance during collaborative operations, thus promoting network efficiency and stability.

Endogenous twin in network elements: By embedding digital twin technology into network element devices, real-time monitoring and analysis of device status, performance, and operating environments can be achieved. Rapid responses to network changes and alerts are enabled. Combined with centralized platforms to integrate twin data from different network elements, this approach supports cross-equipment and cross-regional network collaborative optimization and intelligent decision-making, enhancing the flexibility and scalability of network digital twins.

Business-level digital twin: Exploring user sensing-based digital twins, combining user experience with network behavior and mapping it precisely into the network model. This allows for a better understanding and prediction of the relationship between user needs and network performance. Through user sensing-based data analysis, digital twins can optimize network resource allocation, improve service quality, and dynamically adjust network configurations to meet the needs of different user groups, enhancing user satisfaction.

Cross-domain service and single-domain network management collaboration: Single-domain digital twins focus on data modeling and optimization within a specific domain. Cross-domain digital twins integrate data and models from multiple domains, offering a global perspective and enhancing the collaborative effects between different domains. On this basis, cross-domain collaborative optimization improves resource scheduling, service quality, and multi-level, multi-dimensional intelligent decision-making, providing more accurate predictions, monitoring, and adjustments in resource allocation and service provision within the network, ultimately achieving more efficient, intelligent, and adaptive network management.

Collaborative development with AI LLM technology: Digital twins leverage AI LLMs to provide rich dynamic input data for training and interference in environments closer to real-world conditions, enhancing predictive accuracy and decision-making capabilities. The powerful computing and deep learning capacities of AI LLMs further improve the intelligence of digital twins, enabling them to handle more complex scenarios, predict finer trends, and incorporate more variables and optimization factors into the decision-making process.

2.6 Trend 6: Integration of Digital Intelligence Promotes the Development of Data Elements and Data Infrastructure

Insight and Analysis

The integration of big data and AI is a multifaceted development trend, enhancing data processing, analysis, and application capabilities through intelligent methods. This integration not only optimizes data processing workflows but also strengthens data insights and decision-making support capabilities.

Data elements are the core production factors of the digital economy and have been elevated to a national strategic resource. Data infrastructure serves as the foundational layer supporting all data-related activities, including computing, storage, networks, and software resources to fully tap into the value of data. It represents the evolution of traditional IT infrastructure. The construction of data elements and data infrastructure has become a key factor in digital transformation.

The integration of digital intelligence is the catalyst for the development of data elements and data infrastructure and is accelerating the release of data element value.

The integration of digital intelligence is becoming the key force driving industrial upgrades, technological innovation, and social development. It is the driving force for intelligent upgrades, production method transformation, technological innovation, digital transformation, and the development of one-stop big data platforms.

With their vast data resources and advanced data processing infrastructure, operators are well-positioned to accurately grasp the trend of the integration of digital intelligence, effectively enhancing the intelligence level of their networks and thereby driving business innovation and significant efficiency improvements.

Recommendations

Promote the co-construction of the "AI+" ecosystem: Operators should adhere to open cooperation and mutual benefit, improving the new cooperation mechanism based on complementary functions, healthy interaction, joint efforts, and open sharing. They should work together to expand the "AI+" ecosystem, promote "AI+" infrastructure upgrades, strengthen data governance, enhance the development momentum of digital intelligence services, and create a favorable environment for digital intelligence services.

Data-driven decision-making mechanism: Use big data analysis tools to establish a data-based decision support system. By analyzing user behavior and market trends in real time, operators can quickly respond to market changes, improving the accuracy and timeliness of decision-making.

Deepen data application innovation: Leverage GenAI technology to deeply mine and analyze data, developing innovative applications such as personalized recommendations, targeted marketing, risk forecasting, and intelligent customer service. These applications can enhance user experience and operational efficiency, creating more commercial value for enterprises. Furthermore, Gartner predicts that by 2026, data management tool providers who fail to leverage GenAI will see their annual revenue decline by 10%. GenAI will reduce labor-intensive data management costs by 20% annually, while new use cases will increase threefold.

Data security and privacy protection: Strengthen data security and privacy protection measures to ensure the safe and compliant use of user data. Through data desensitization and encryption technologies, operators can protect user privacy and build user trust.

Talent development and skill enhancement: Invest in employee training in big data and AI skills to improve the team's data literacy and analytical capabilities. By fostering cross-department collaboration, operators can cultivate multi-disciplinary talent with expertise in data analysis and business understanding.

Data sharing and collaboration: Establish a data-sharing mechanism and jointly develop new data products and services with partners. By opening data interfaces, operators can promote data circulation and value creation.

Continuous technological innovation: Continuously track and invest in the latest technologies in the fields of big data and AI, leveraging operators' data advantages to maintain technological leadership.

By implementing the above recommendations, operators can better leverage big data and AI technologies to enhance service quality, optimize operational efficiency, and maintain a competitive edge in the market.

2.7 Trend 7: Evolution of Multi-Agent Collective Intelligence Collaboration

Insight and Analysis

We are entering an era where AI reconstructs everything, driven by LLMs that propel intelligent transformation.

AI-centered intelligent architecture: The evolution of new product system architectures driven by AI is a key stage in the next decade toward AI-powered intelligent architecture. The new intelligent architecture needs to be built around LLMs, breaking traditional limitations.

Multi-agent collaboration as the key technical direction for solving complex problems in the future: As the intelligence of LLMs increases, multi-agent collaboration based on LLMs will be the key technical direction for addressing the complex challenges of autonomous networks. Agents, due to their inherent intelligence capabilities, such as autonomous decision-making, learning adaptation, and environmental interaction, can solve complex problems by sensing the environment, analyzing information, and applying algorithms to form and execute policies. Furthermore, agents possess self-optimizing capabilities, continuously improving their performance over time to adapt to changing environments and task requirements. More importantly, agents are highly flexible and goal-oriented, and in the future, they will become more efficient and effective in handling complex problems and seamlessly interacting.

New paradigm for autonomous network design based on multi-agent collaboration: The evolution of AI-driven autonomous networks requires starting with the ultimate goal of intelligence, thinking backward, and designing today's products. In the future, communication networks based on agents will be better able to conduct dynamic resource scheduling, fault prediction and repair, and user experience optimization, thus creating significant commercial value and shaping future competitive advantages.

Currently, the application of agents in autonomous networks is transitioning from innovative exploration to practical implementation. The five AI paradigms and three technical engines we propose are critical judgments and technical supports for the implementation of agent applications in autonomous networks.

Key design elements for agent applications: The AI-centered autonomous networks of the future will bring entirely new product forms.

- User-adaptive experience: AI can provide personalized preferences based on user likes, interests, and business execution logic, continuously learning individual or business feature data to offer highly adaptive services.
- The design of agent behaviors, collaboration mechanisms and processes between humans and agents, as well as the expression of awareness and outcomes, will become key focuses of design. The design of agent behavior needs to take into account multimodal heuristic dialogues, internal and external data sensing, proactive handling and suggestions, and, around autonomous iteration, dynamically generate and present personalized multimodal results tailored to different scenarios.

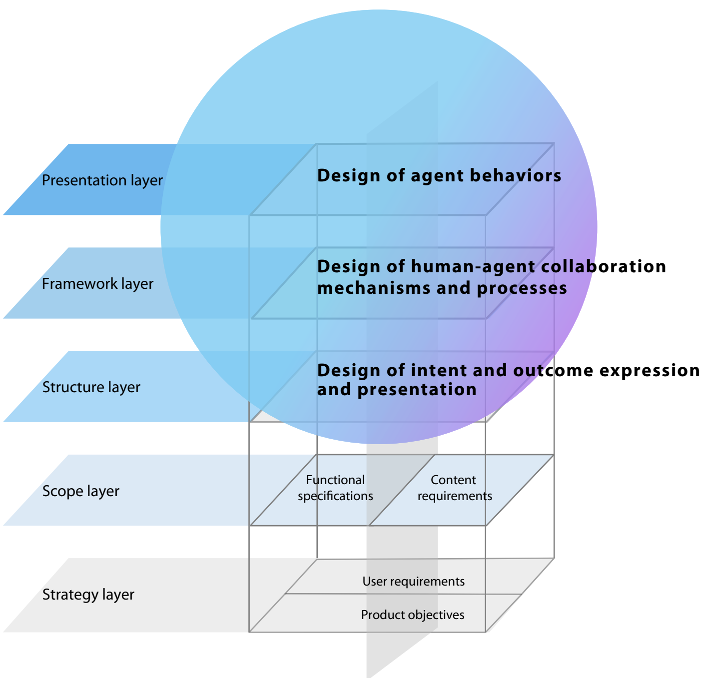


Figure 3. Design Framework for Agent Behavior

Multi-agent collective intelligence collaboration enables rapid end-to-end automation of business processes: As an advanced technological paradigm, multi-agent collective intelligence collaboration involves multiple independent and autonomous agents working seamlessly together within the network. These agents share information, learn from each other's behavior patterns, and adapt to one another, achieving full automation of business processes. The core of this collaborative mechanism lies in its capacity to quickly respond and self-optimize. When business requirements change, each agent can immediately adjust its policy, collectively accomplishing tasks and driving business processes forward through real-time learning and decision-making. They are capable of automatically identifying issues, analyzing data, executing tasks, and providing continuous feedback and iteration throughout the process, forming an end-to-end closed-loop. Multi-agent collective intelligence collaboration not only simplifies business management but also enhances a company's competitiveness and innovation, becoming a key driver for transforming modern business operations toward greater intelligence and efficiency.

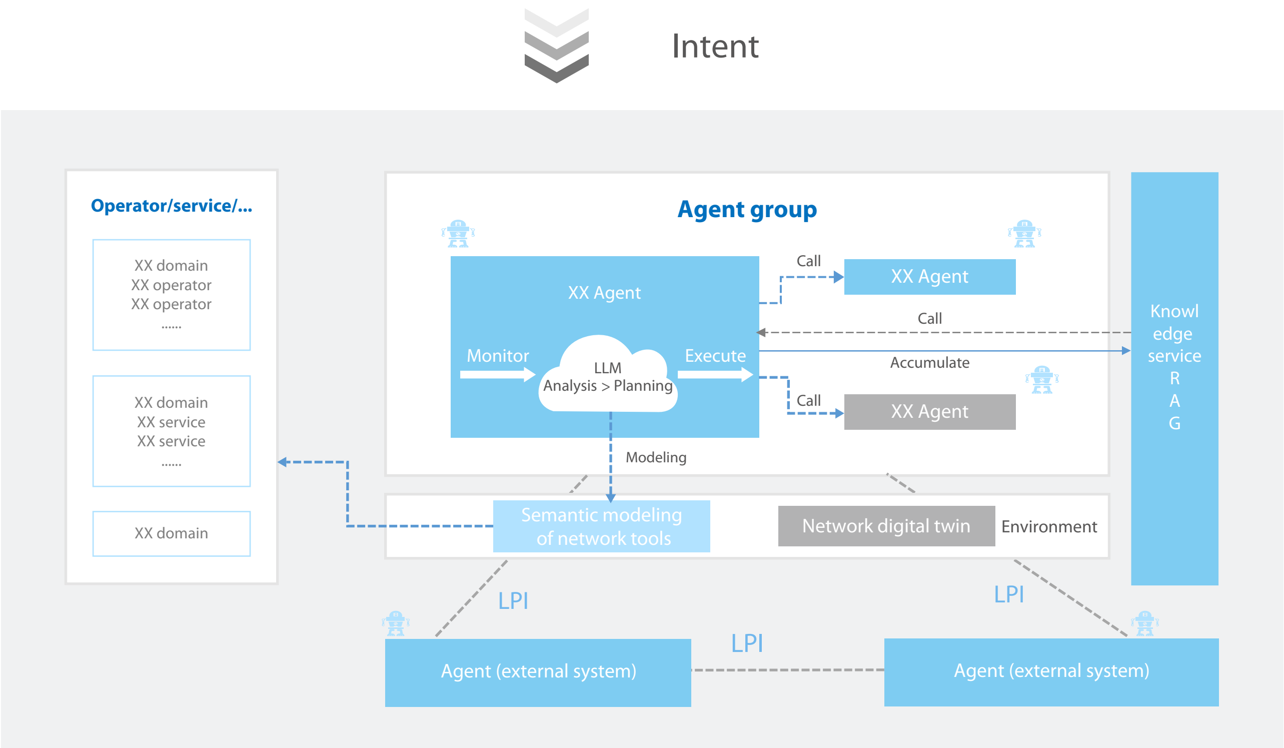


Figure 4. Multi-Agent Collective Intelligence Collaboration

Future challenges in the evolution of multi-agent collaborative networks :

Agent behavior control : Agent behavior control is entering an era of mass production, with agents exhibiting three key attributes: first, the ability to act independently based on specific goals, demonstrating autonomous decision-making; second, the capability to interact seamlessly with the external environment and effortlessly utilize various software tools; and third, the potential for indefinite operation with evolutionary capabilities, allowing agents to gradually self-optimize through feedback during their workflows, such as acquiring new skills and refining existing skill combinations.

The operational logic of agents may lead to harmful deviations during the pursuit of specific objectives. In some cases, agents may only grasp part of the literal meaning of a goal without fully understanding its true intent, resulting in abnormal behavior when processing certain objectives. When agents make decisions based on environmental sensing and issue commands to humans in the real world, the potential for harmful commands being mistakenly executed increases, particularly in situations where humans cannot adequately assess the outcomes.

Some agents lack effective deactivation mechanisms after being created. These agents may continue to operate in environments entirely different from their original deployment conditions, diverging significantly from their initial purpose. Additionally, agents might interact in unforeseen ways, causing unexpected incidents.

Standardization research in the field of multi-agent: Currently, the field of multi-agent systems is in its nascent phase, lacking a unified technical framework and standardized protocols. Researchers are actively exploring and striving to establish a set of universal guidelines to overcome compatibility challenges between different agent systems, ensure effective interoperability of information, and improve the precision of cross-system semantic understanding. This is an ongoing process that requires continuous effort to identify best practices, laying a solid foundation for the broad integration and collaboration of multi-agent systems.

Recommendations

To address the emerging trends and challenges of new AI agent applications, it is essential to strengthen the agent-based AI application architecture paradigm, ensuring secure agent operations and standardized communication capabilities. The recommendations are as follows:

The establishment of suitable agent application architecture paradigms

Based on various Level 0 LLMs, continue to explore and research agent-based application architecture paradigms in dimensions such as human-computer interaction, agent collaboration, and agent openness services.

Agent application operational security

Considering the autonomous evolution features of agents, it is necessary to filter harmful knowledge, enforce behavioral norms for autonomous execution, and establish operational control mechanisms such as electronic fences for executing agents.

Agent service-oriented openness capabilities

Explore and establish standardized communication protocol specifications for both homogeneous and heterogeneous agents. It is necessary to collaborate with ecosystems to formulate unified agent communication protocols and service-oriented openness capabilities.

Engineering production and application of agents

Use AI to produce AI, and leverage one-stop tool-based engineering capabilities to achieve efficient development and delivery, supporting the recycling and closed-loop iteration of knowledge and data.

03

ZTE Highly Autonomous Network Evolution Solution

3.1 Overall Approach to High Autonomy

"Zero-X Experience and Self-X O&M" represents the vision and goals of autonomous networks, such as providing customers with the "three-zero" experience of "zero waiting, zero fault and zero contact", and building the "three-self" capabilities of "self-configuration, self-repairing and self-optimization" for O&M. To align with these goals, TM Forum has developed a six-phase evolution path (Level 0-Level 5), where the Level 4 autonomous network represents a significant shift from traditional human-directed to network-directed automation.

Focusing on high-value scenarios, the goal is to build automated and intelligent operation capabilities across the entire network lifecycle, yielding measurable benefits such as improved quality, increased revenue, cost reduction, and enhanced efficiency. These are the remarkable features of Level 4 autonomous networks, which present unique challenges across various operators' networks.

The AIR Net solution is a high autonomous network evolution solution proposed by ZTE for global operators. It is characterized with:

- The open decoupling of data and capabilities to enable end-to-end automation processes;
- Key intelligent technologies such as the collaboration of large and small models, and digital twins, used to lower the skills threshold for network operations, gradually replacing manual work to realize autonomous closed-loop of the network and solve O&M problems;
- A modular architecture, for rapid orchestration and deployment of autonomous capabilities for various scenarios;
- Capabilities to define value scenarios and outcomes, to facilitate the closure of the business performance loop.

3.2 Selecting High-Value Scenarios for Level 4 Highly Autonomous Networks

In June 2024, at the Global Digital Transformation Summit (DTW24), TM Forum, in collaboration with China Mobile, Vodafone, Telefonica, and other industry partners, released 《Autonomous Networks Level 4 Industry Blueprint – High-Value Scenarios》. The white paper outlines a phased evolution path toward Level 4 highly autonomous networks focused on high-value scenarios, providing telecom operators with a reference framework for progressing to Level 4 autonomy.

End-to-end high-value scenarios are critical carriers and focus areas for achieving Level 4 autonomous networks. Defining these scenarios helps delineate the practical scope of autonomous networks, enable potential business or operational improvements, enhance return on investment, and ultimately establish a commercial closed loop. By identifying these scenarios and leveraging network open decoupling empowerment, along with the transition from +AI assistance to AI+network, breaks and bottlenecks in end-to-end processes will be eliminated, enabling the network to progress from automation and intelligence in sub-segments to a fully closed-loop, value-driven end-to-end system.

TM Forum has organized the first phase of high-value scenarios from two dimensions: business-oriented and network-oriented. In the business dimension, scenarios are categorized based on three major business types: personal, home, and government-enterprise, and are further refined across three dimensions: business provisioning, network complaint processing, and business quality optimization. In the network dimension, scenarios are analyzed across various domains, including wireless networks, core networks, network clouds, transmission networks, IP networks, and business platforms, based on dimensions such as equipment network access, network fault monitoring, network change monitoring, network performance optimization, network energy efficiency optimization, and resource utilization optimization. Based on the operational value and technical maturity, ZTE has expanded the 20 typical scenarios recommended by TM Forum and selected the following 29 high-value scenarios to support:

<table>
<tr><th colspan="2" rowspan="2">Targeted Business</th><th colspan="3">Personal Business</th><th colspan="3">Home Business</th><th colspan="7">Government & Enterprise Business</th></tr>
<tr><th>Voice</th><th>Internet access</th><th>…</th><th>Household broadband</th><th>Internet TV</th><th>…</th><th>5G private network</th><th>IoT</th><th>Cloud access dedicated line</th><th>Cloud interconnect dedicated line</th><th>Internet dedicated line</th><th>Data dedicated line</th><th>…</th></tr>
<tr><td rowspan="3">Operations [DROP]</td><td>Business provisioning</td><td></td><td></td><td></td><td>▲</td><td></td><td></td><td>▲</td><td></td><td>▲</td><td>▲</td><td>▲</td><td>▲</td><td></td></tr>
<tr><td>Network complaint processing</td><td colspan="3">▲</td><td>▲</td><td></td><td></td><td>▲</td><td>▲</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Business quality optimization</td><td colspan="3"></td><td>▲</td><td></td><td></td><td colspan="2">▲</td><td></td><td></td><td>▲</td><td></td><td></td></tr>
</table>

<table>
<tr><th colspan="2">Network-Oriented</th><th colspan="2">Wireless network (4G, 5G)</th><th>Core network (EPC, 5GC, etc.)</th><th>Network cloud (Server, Virtual Layer)</th><th>Transmission network (OTN, SPN, PON, etc.)</th><th>IP network (CMNET, cloud private network, etc.)</th><th colspan="2">…</th></tr>
<tr><td>Planning and construction</td><td>Equipment network access management</td><td colspan="2"></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td rowspan="2">Maintenance</td><td>Network fault monitoring</td><td colspan="2">▲</td><td>▲</td><td>▲</td><td>▲</td><td>▲</td><td></td><td></td></tr>
<tr><td>Network change monitoring</td><td colspan="2"></td><td>▲</td><td>▲</td><td>▲</td><td>▲</td><td></td><td></td></tr>
<tr><td rowspan="3">Optimization</td><td>Network performance optimization</td><td>▲</td><td>▲</td><td></td><td></td><td></td><td>▲</td><td></td><td></td></tr>
<tr><td>Network energy efficiency optimization</td><td>▲</td><td>▲</td><td></td><td></td><td>▲</td><td></td><td></td><td></td></tr>
<tr><td>Utilization optimization</td><td colspan="2"></td><td></td><td></td><td>▲</td><td></td><td></td><td></td></tr>
</table>

Figure 5. ZTE Recommended Level 4 High-Value Scenarios (Phase 1)

3.3 High Autonomous Network Solutions and Application Paradigm Evolution

3.3.1 Solutions

As a promoter and practitioner in the autonomous network industry, ZTE, leveraging its longstanding technical expertise and extensive experience in autonomous network implementation, continues to explore and innovate. Following the principle of value and effectiveness-driven, ZTE has introduced the AIR Net high autonomous network evolution solution. By leveraging open decoupling and digital intelligence engines, the solution supports end-to-end automated and intelligent O&M and operations. The solution is structured across three levels: cross-domain collaboration, single-domain autonomy, and endogenous intelligence, providing hierarchical, domain-based, and tiered evolutionary capabilities. It can be deployed independently or in the cloud, empowering digital network operations.

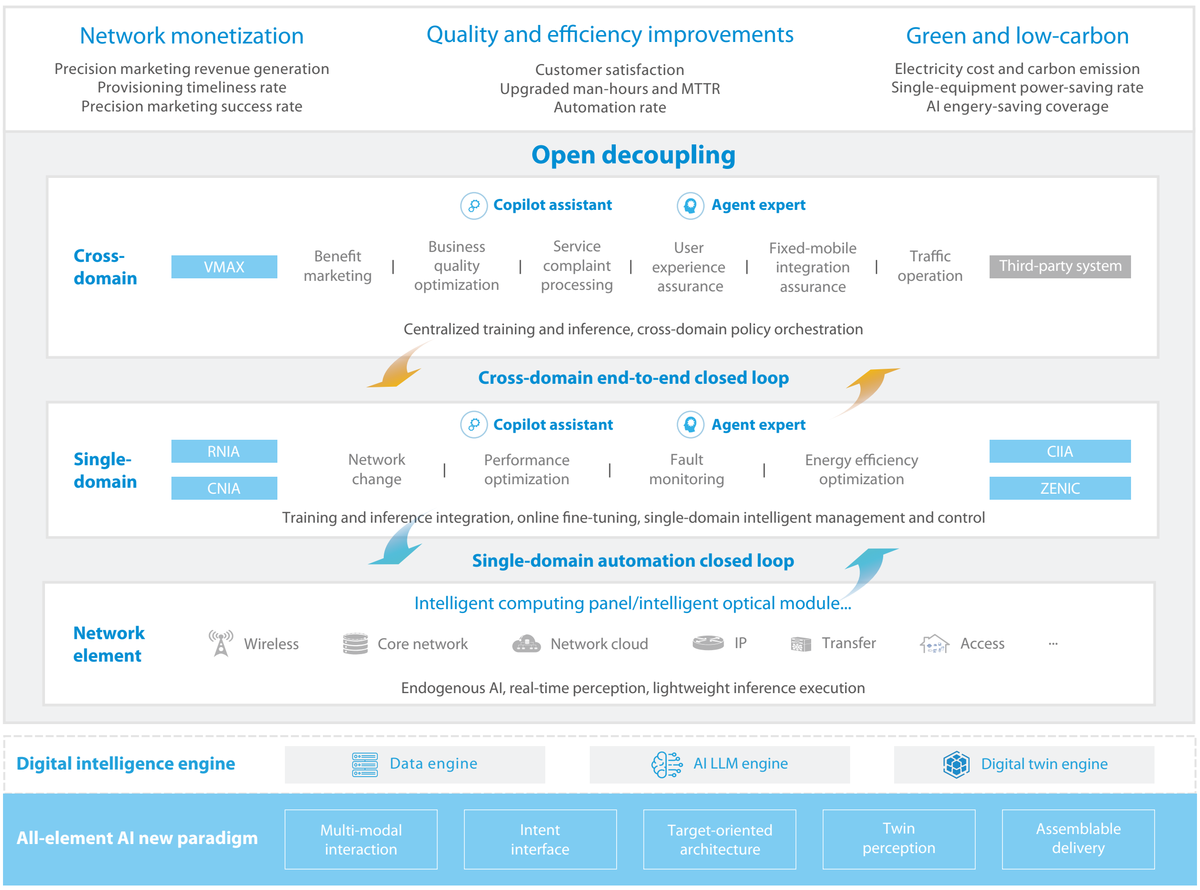


Figure 6. Panorama of ZTE Autonomous Network Solution

The solution is built on a All-element AI new paradigm to construct a digital intelligence engine, focusing on value scenarios to achieve the core goals of network monetization, network quality improvement, O&M efficiency improvement, and green low-carbon autonomous networks. Brief descriptions are as follows:

1. Based on the iterative evolution of five paradigms, the digital intelligence engine has been designed:

- The data engine is responsible for data collection, governance, storage, and processing;
- The AI LLM engine drives intelligent applications through the LLM training and inference toolchain, large and small model collaboration, model libraries, and agent development frameworks;
- The digital twin engine realizes virtual-physical interaction, decision analysis, and visualization. Together, these three engines establish an intelligent network O&M system, propelling autonomous networks toward higher levels of intelligence.

2. The endogenous intelligence of network elements is the foundation for realizing high autonomy. By introducing real-time sensing equipment and AI inference capabilities into cloud network equipment across wireless, core networks, network clouds, transmission, IP, and access layers, the solution realizes various agents of network elements, supporting real-time network optimization and experience improvement.

3. The OMC intelligence for single-domain autonomy is based on the original network element management, supporting the implementation of automation closed loops through data openness, capability openness, and decoupling. It further integrates the training and inference integration and online fine-tuning digital intelligence engine, provides Copilot Assistant (role-oriented, AI-assisted agent applications) and Agent Expert (scenario-oriented, AI-driven agent applications), empowering time and manpower-intensive process phases. This enables the automation and intelligent closed-loop of value scenarios such as business provisioning, change control, performance optimization, fault management, and energy efficiency optimization across wireless, core networks, network clouds, transmission networks, IP networks, and access networks.

4. The solution enables cross-domain collaborative intelligence. Through centralized training and cross-domain policy orchestration, ZTE provides Copilot Assistant and Agent Expert capabilities, empowering end-to-end automation. This ensures end-to-end cross-domain intelligent automation closed loops for scenarios such as service quality assurance, poor-quality optimization, fault monitoring and processing, and user complaint processing.

5. The solution accelerates the elimination of bottlenecks and gaps through data openness, capability openness, and decoupling. Additionally, it supports manufacturer capabilities for orchestration and invocation by upper-layer systems, driving the realization of end-to-end automated and intelligent O&M and operations.

6. In the solution, LLM capabilities are fully leveraged, providing application as follows:

COPILOT ASSISTANTS(6 CATEGORIES, 20 PRODUCTS)

PRODUCT NAME	USER ROLE
Fault assistant	Monitoring engineer
Household broadband installation and maintenance assistant	Installation and maintenance engineer
Monitoring assistant	O&M engineer
Network observation assistant	O&M engineer
Q&A assistant	O&M engineer
Core network O&M assistant	O&M engineer

AGENT EXPERTS (7 CATEGORIES, 21 PRODUCTS)

PRODUCT NAME	APPLICATION SCENARIO
Network change expert	Network change
Network optimization expert	Network optimization
Fault monitoring expert	Fault monitoring
Complaint analysis expert	Complaint processing
Energy efficiency optimization expert	Energy efficiency optimization
Assurance expert	Key event assurance
Network observation expert	Network planning

Figure 7. LLM Application in ZTE Autonomous Networks

The AIR Net solution provides operators with three core values: network monetization, quality and efficiency enhancement, and green and low-carbon.

Network monetization

Through in-depth insights into network data, the solution enables precise marketing and differentiated guarantees for influencer live streaming, turning network advantages into market advantages and enhancing the profitability of the network. For example, in the precise marketing scenario, data from the influencer live-streaming analysis system is used for precise marketing, resulting in an overall improvement of 16% compared to traditional offline promotions and 5% to online promotions.

Quality and efficiency enhancement

Through intelligent computing single boards, the solution proactively identifies poor-quality users, improves the accuracy of end-to-end demarcation and positioning, and facilitates timely remediation and self-optimization, improving complaint processing efficiency and customer satisfaction. For example, in the mobile service complaint processing scenario, the demarcation accuracy increases by 85%, the front-line maintenance dispatch decreases by 10%, and customer satisfaction improves by 5%.

Green and low-carbon

Through the sinking of AI capabilities and the dual-layer intelligent collaboration between network elements and the network, the solution enables the automated formulation of optimal energy-saving policies, further increasing power-saving rates and contributing to green and low-carbon development. For example, in the wireless 5G energy efficiency optimization scenario in a city, a 15% energy-saving rate per equipment is achieved, saving 5.33 million kWh annually and reducing electricity costs by RMB 4.42 million (in a prefecture-level city of a province).

3.3.2 Technical Architecture

The technical architecture of the AIR Net autonomous network solution is shown in the figure below, relying on the digital intelligence engine and autonomous services to support the design, development, and deployment of intelligent applications.

The autonomous service is responsible for aggregating the autonomous network service capabilities in various fields, such as business provisioning services, complaint processing services, quality optimization services, fault processing services, change monitoring services, performance optimization services, energy efficiency optimization services, equipment network access services, utilization optimization services, low-code orchestration services, etc., supporting the use of the autonomous network application layer.

The digital intelligence engine is mainly composed of three technical engines: data engine, AI LLM engine, and digital twin engine. These three engines collaborate to provide mutual support. Among them, the data engine is the foundation, which provides the required data resources for the AI LLM engine and the digital twin engine and ensures the effective implementation of data governance. The AI LLM engine is the intelligent center of the future network, supporting the construction of digital twin models by training LLMs in various scenarios. At the same time, the twin engine in the digital twin platform generates new data, which can be used to further train and optimize the LLMs. This mutually empowering relationship ensures that the three technological engines work closely together, driving technological innovation and development. In short, the data engine provides core resources; the AI LLM engine serves as the intelligent center for processing resources intelligently; and the digital twin engine applies the processed information to actual scenarios, forming a closed-loop, efficient innovation ecosystem.

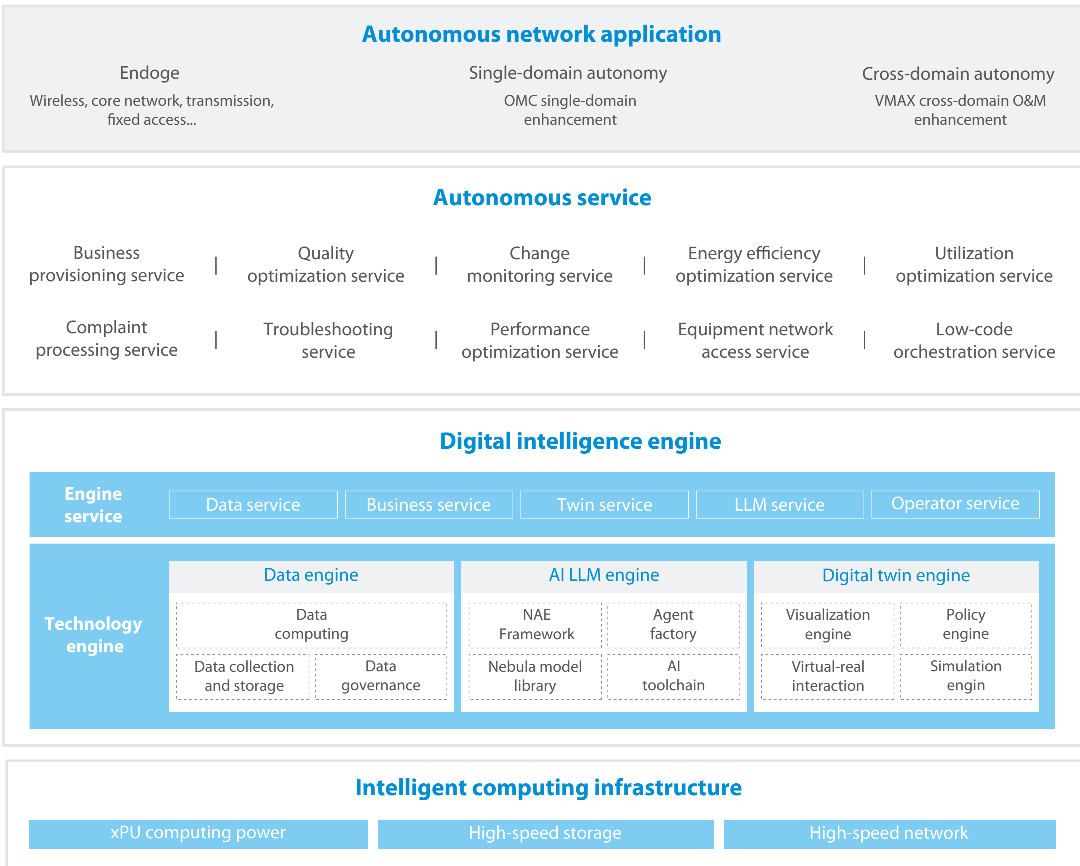


Figure 8. ZTE AIR Net Technical Architecture Diagram

Data engine

The data engine supports intelligent applications through functions such as data collection, data governance, data storage, and data processing. It supports multi-source collection and format processing, various data sources, and structured, semi-structured, and unstructured data. It provides data conversion (e.g., format, encoding) and cleansing (e.g., deduplication, processing missing values), supports integrated data storage (lakehouse architecture), ensures data security and privacy protection, and provides data encryption, data access control, data interfaces, data sharing, and data interaction. In the future, GenAI capabilities will play a crucial role in the data engine, reducing manpower-intensive data management costs by 20% annually. New use cases will increase threefold. LLMs in the data engine will take the form of data governance agents, and the development of AI4Data will require the engineering capabilities of agent + tools (APIs) for data decision and process orchestration applications.

AI LLM engine

The AI LLM engine plays an important role in model training, reuse, and agent development through the LLM training and inference toolchain, LLM library, and agent engine. It serves as the intelligent center of the autonomous network and enables the widespread application of GenAI technologies within the autonomous network.

The LLM training and inference toolchain integrates an efficient parallel computing framework during the training phase, providing a series of optimization methods for model training, including hyperparameter adjustment tools, model evaluation tools, and more. Developers can use these tools to evaluate model performance comprehensively. During the inference phase, the LLM training and inference toolchain optimizes the model to adapt to different hardware environments. It also uses quantization techniques to convert model parameters from high-precision data types to low-precision data types, reducing storage requirements and computing amounts, and improving inference speed.

The LLM model library provides model reuse and sharing. It stores various pre-trained domain-based basic LLMs. By fine-tuning these models on specific datasets, a high-performance model for specific scenarios can be quickly obtained.

The agent engine provides a framework for building and customizing agents, supporting the agent factory to provide a convenient development environment for constructing application agents. By defining the agent's goals, behaviors, and interaction rules, agents with specific functions can be developed. Different agents can communicate and collaborate to complete complex tasks.

Digital twin engine

The digital twin engine achieves comprehensive mapping, analysis, and optimization management of physical entities through the coordination of multiple built-in sub-engines.

The bidirectional data flow, status synchronization, and mapping between physical entities and digital twin models enable virtual-physical interaction, allowing the digital model to reflect the status of the physical entity in real-time. The simulation engine can simulate the system behavior of digital twin objects and predict the system's operational status and development trends under different input conditions and parameter settings. It is used for testing and validating various scenarios and evaluating the feasibility and effectiveness of policies. The decision engine analyzes data from digital twin models and other relevant sources, evaluates the potential impact of each solution, and generates the optimal decision solution. The visualization engine presents complex data of digital twins in intuitive forms such as graphics, images, and animations, allowing users to quickly understand the data's meaning. Users can interact with the digital twin model through the visual interface and observe the effect after adjustment in real-time, enhancing interactivity and engagement during the decision process.

3.3.3 AI Agent Application Paradigm Evolution

Modern AI agent applications, such as role-assisting Copilot Assistants and scenario-intelligent Agent Experts, harness digital intelligence engines—including the Data Engine, AI LLM (Large Language Model) Engine, and Digital Twin (DT) Engine—to deliver advanced functionality, adaptive problem-solving, and support for complex scenarios like:

- Network fault diagnosis and recovery.
- Intelligent O&M (Operations and Maintenance) knowledge Q&A.
- Network performance optimization.
- Energy efficiency management.

These AI-driven agents are poised to redefine autonomous network workflows, marking a paradigm shift from traditional O&M tools in interaction methods, interface design, system architecture, business capabilities, and development/delivery processes. This transformation stems from the evolution of the AI Agent Application Paradigm, detailed below:

Interaction Paradigm

Definition : Human-computer interaction transitions from GUI (Graphical User Interface) → LUI (Language User Interface) → MUI (Multimodal User Interface).

Features : Intent-centric, multimodal contextual dialogue (text, voice, vision).

Proactive data sensing, autonomous processing, and intelligent recommendations.

Dynamic iteration based on user intent, generating personalized, scenario-aware multimodal outputs.

Action Plan : Build on existing GUI frameworks.

Integrate LUI, MUI, and GeneUI (Generative User Interface) to augment or replace legacy UI components.

Enable coexistence and interoperability of multiple interaction modalities.

Interface Paradigm

Definition : The interface paradigm evolves from traditional APIs to intent-driven interaction interfaces between users and systems and systems and systems.

Features : Users define only their desired goals or expectations (not implementation details).

The system autonomously infers, analyzes, and automates execution paths to fulfill these objectives.

Action Plan : Use APIs as the foundational layer.

Achieve >90% accuracy for LPI (Language-Programming Interface) recognition and parameter extraction.

Transition LPI to mainstream adoption and enable cross-system LPI interoperability.

Architecture Paradigm

Definition : The product architecture shifts from process-centric to goal-centric design.

Features : Agent-driven architecture.

Dynamic, generalized, and composable product forms with unbounded goal-task adaptability.

Action Plan : Implement Agent-as-a-Service (AaaS) with decoupled, modular components.

Adopt a multi-agent design framework (covering agent teams, roles, skills, collaboration, communication, exception handling, and observability).

Twin Perception Paradigm

Definition : Transition from local visibility to full-link data-driven digital twins, enabling virtual reflection and control of physical networks.

Features : Achieve perceptibility, visibility, trustworthiness, and controllability across infrastructure, applications, and business layers (including prediction, analysis, and decision-making).

Action Plan : Enable single-layer local visibility.

Establish cross-layer MTL (Multi-Tier Link) full-link visibility.

Implement end-to-end business process awareness and digital twin synchronization.

Delivery Paradigm

Definition : Shift from version/product-driven R&D delivery to goal-driven, orchestrated assembly-based engineering delivery.

Feature : Git-centric workflows (branching, versioning).

Kubernetes CRD-aligned cloud-native deployment with declarative manifests and composable delivery.

Action Plan : Integrate Git, CRD, and Operators to enable CNF (Cloud-Native Function) and EM (Edge Module) versions.

Support online testing, changes, instantiation, and upgrades.

By applying these paradigms to the AIR Net autonomous network solution, AI-agent-based applications will undergo transformative evolution, enabling adaptive, goal-oriented intelligence across systems.

04

ZTE Highly Autonomous Network Practice Cases

4.1 Core Network Upgrade Process Automation

Case Background

With the rapid evolution of emerging technologies and services, the frequency of iterations and updates in core network functions increases significantly. However traditional upgrade method face several challenges, including high risks of manual operation error, uncontrollable upgrade progress, and inefficient cross-organizational collaboration, which then impact on the delivery efficiency of new functions. Establishing an end-to-end, full-process automated upgrade system and improving the security and reliability of upgrade operations have become common concerns in the industry.

Solution

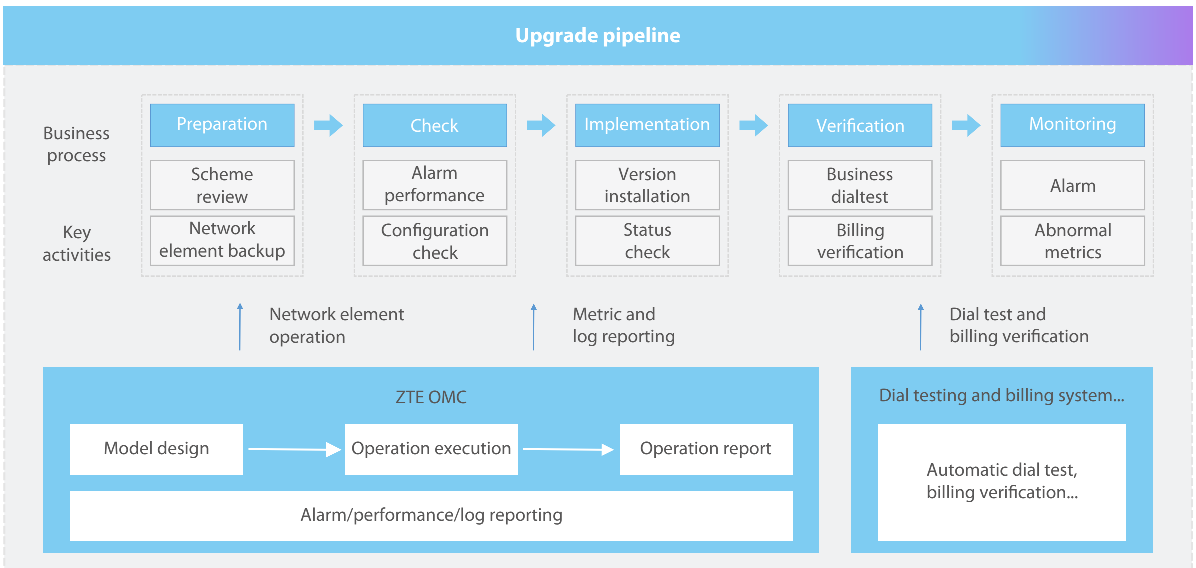


Figure 9. Core Network Automated Upgrade Process Diagram

ZTE core network upgrade automation solution integrates with the OMC system and uses IT continuous integration and deployment (CI/CD) practices. The solution introduces a model-driven network-element automated upgrade mode, dividing the network element upgrade process into two phases: model design and model execution.

Model Design Phase, Leveraging a model design engine, the solution offers modular, atomic capabilities tailored to various network change scenarios. These atomic capabilities are combined with specific NE upgrade requirements to rapidly develop automated upgrade models. Once validated, these models are published to a version repository in alignment with NE version releases. Currently, the solution provides comprehensive coverage for 5GC and IMS network elements. Future expansions will include scenarios such as capacity scaling and configuration changes, further enhancing operational efficiency across all network change scenarios.

Model Execution Phase, The model execution engine ensures a visualized, controllable, and manageable execution environment. With intuitive dashboards, operators can monitor upgrade progress in real time and handle anomalies within one-click, significantly reducing workload. The solution fundamentally eliminates human error through complex operations automation and intent-based workflows.

By enabling the decoupling between system data and capabilities, the OMC integrates seamlessly with operators' O&M systems to build an end-to-end automated production line. This production line supports intelligent orchestration of critical functions such as service dial tests, billing verification, and upgrade validation, driving the transition from localized automation to full-process, closed-loop automation.

Results

ZTE automated upgrade solution has been successfully deployed in both domestic and international markets. The solution achieves a 95% full-process automation rate for NE upgrades, saving an average of 70% in man-hours per upgrade while ensuring near-zero disruption to live network operations. By eliminating manual interventions, such as frontline script modifications, the solution brings human errors down to nearly zero, ensuring highly reliable and efficient upgrade processes.

4.2 Intelligent Handling of Network Cloud Faults

Case Background

With the convergence of cloud and network, the network cloud infrastructure has become the cornerstone of the entire communication network. It is crucial to ensure its high-quality and stable operation. However, the O&M work of network cloud is more complex than that of traditional networks. It involves vast resource pools, including physical hardware as well as virtualized compute, storage, and network resources, making daily O&M tasks highly challenging.

Traditional fault detection systems primarily rely on customer complaints and alarm information, often resulting in reactive and delayed fault detection. Once a fault occurs, the root cause locating process can take a long time and require the involvement of many experts, increasing the risk of slow recovery from major service faults.

Using AI and big data, the solution proposes both to proactively predict and prevent faults, and to quickly recover once an issue occurs.

Solution

CIIA, ZTE's cloud network O&M product, leverages Nebula LLMs and agents/RAG enhancement technologies to provide intelligent analysis, content generation, and daily O&M report generation capabilities based on resource pool object alarms, performance indicator, logs, and other data, empowering cloud network O&M application practices.

Network cloud fault handling is based on multi-agent intelligent O&M and uses the following key technologies:

- A general language model is fine-tuned (FT) and supervised fine-tuned (SFT) with a mixture of domain-specific corpus and collected inference corpus to create a domain-specific LLM, improving the accuracy rate of downstream tasks by 6%.
- The cutting-edge multi-agent technology can complete more than 70 times of dynamic task decomposition with an accuracy rate of over 90%. It represents an industry-first solution for intelligent, accurate completion and deployment of complex domain-specific tasks.
- The domain-specific LLM includes a log LLM that can automatically analyze abnormal log sequences and perform diagnostics and a plug-and-play time-series LLM. All evaluation metrics outperform industry benchmarks.

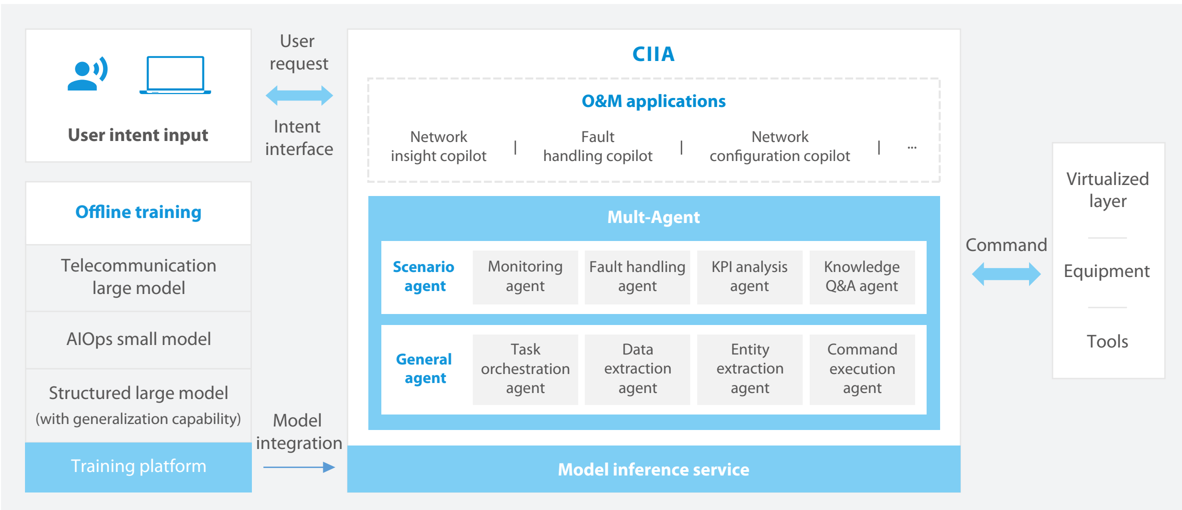


Figure 10. Network Cloud Faults Intelligent Handling Solution Diagram

Results

The CIIA product is already deployed in Chinese market in several regions, reducing reliance on experts and significantly saving manpower. The time for identifying and diagnosing switch faults has reduced by 85%, from over 140 minutes to under 20 minutes.

4.3 Intelligent Customer Complaint Handling of Mobile Network Service

Case Background

With the rapid spread and booming development of mobile communication services, major operators have invested heavily to build large-scale, technologically advanced infrastructure networks, aiming to provide users with all-round basic communication and mobile Internet services, including high-definition voice calls, smooth streaming, popular short video browsing, and various over-the-top (OTT) applications. At the same time, user expectations for network service quality have risen sharply, with a focus not only on speed but also on ensuring service stability and personalized needs. In a fiercely competitive market, operators face unprecedented challenges due to the increasingly complex network architecture and diversified business demands. Ensuring optimal user experience despite cost constraints, while efficiently dealing with customer complaints when they occur, are key priorities for operators.

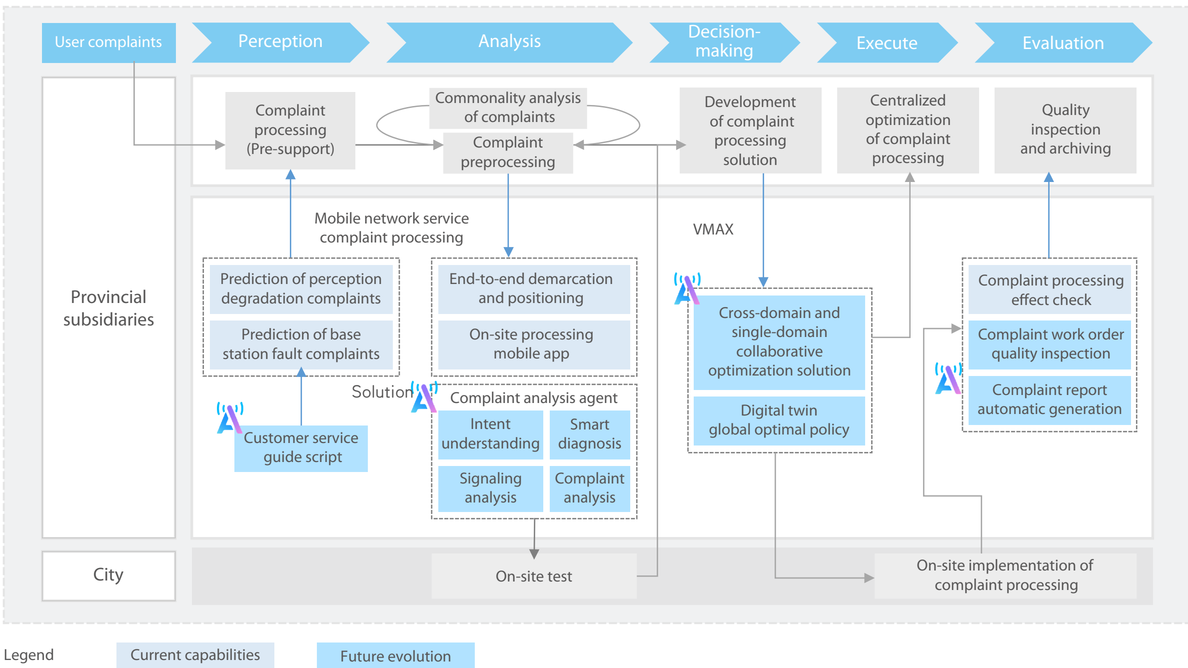


Figure 11. Mobile Network Service Customer Complaint Handling Diagram

Solution

ZTE's solution features Deep Packet Inspection (DPI) probes at key communication interfaces in the mobile network to achieve real-time, comprehensive data collection of mobile network. By leveraging service identification algorithms and perception metric modeling techniques, it can accurately reflect and model user network behaviors. Adding desensitized user profiles, real-time data processing, and in-depth network issue analysis, the tool evaluates user service perception issues, and quickly locates the root causes of issues such as radio network coverage, network quality, and server faults.

Results

Compared to traditional complaint handling methods, the intelligent complaint handling effectively reduced customer complaints by 20% through earlier intervention and greatly shortens the response time for complaint handling. Overall, the mobile service customer complaint handling time is reduced by 50%, the overall demarcation accuracy rate reaches 84%, and the workload of field maintenance is reduced by 10%.

4.4 Service Quality Optimization for Home Broadband Service

Case Background

With the rapid growth of FTTH (Fiber to the Home) , the focus of user concerns has shifted from bandwidth improvement to service experience. Currently, most of the faults can only be reactively identified following user complaint, with 90% of complaints relating to poor service quality. Existing network optimization methods primarily rely on network KPI, lacking systems to measure and monitor user experience. It is then necessary to introduce user experience metrics and proactively identify home broadband users with poor service quality, accurately demarcate and locate the root causes of quality issues, and optimize service quality ahead of user complaints to improve customer satisfaction.

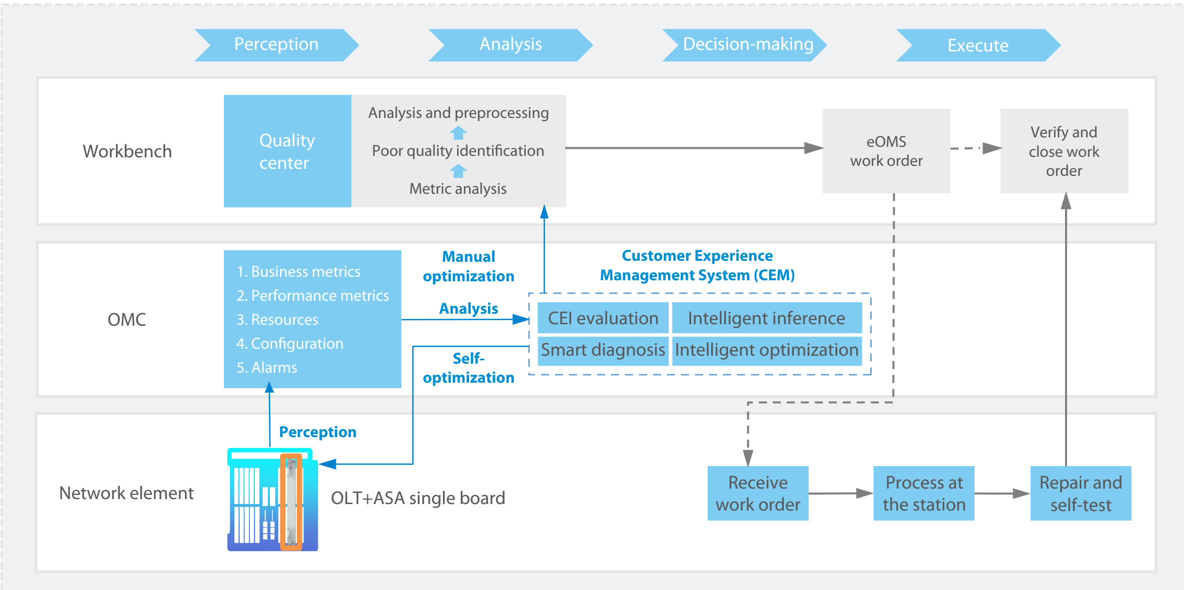


Figure 12. Service Quality Optimization for Home Broadband Service

Solution

The service quality optimization solution for the home broadband service leverages the User Experience Management System (CEM) with an innovative architecture combining Telemetry technology and an ASA intelligent module integrated into the OLT (Optical Line Terminal). The ASA intelligent module embedded in the OLT processes user data collection, preprocessing, compression, data analysis, and closed-loop decision issuance. After the ASA intelligent module identifies and analyzes user business data, it reports user business flow data using the Telemetry protocol, reducing the data reporting interval to 1 minute. Customer Experience Management System (CEM) rapidly identifies end-to-end service quality issues, demarcates and locates root causes, and supports operators in comprehensively and efficiently improving network quality and optimizing user experience.

This solution performs comprehensive home broadband service analysis, covering 6 major service categories and 18,000+ content sources. Using self-developed algorithms, it quantifies user experience and provides 24/7 call detail record backtrack analysis, enabling visualization of the Customer Experience Index (CEI). The end-to-end intelligent demarcation and location of poor service quality for home broadband users leverage machine learning algorithms for anomaly detection and data cleansing. By integrating supervised learning model, convolutional neural network (CNN), and other AI algorithms, the solution performs precise analysis of home broadband service flow data. It generates a list of poor service quality home broadband users and fault root causes, supporting the optimization of home broadband service quality and effectively enhancing operator service performance.

Results

The solution rapidly and accurately demarcates and locates the root causes of poor service quality in end-to-end home broadband services, reducing the demarcation and location time to less than 10 minutes. Together with a Chinese operator in [XX] province, the solution achieved an accuracy rate of over 85% for poor-quality service identification and for root cause location, significantly improving O&M efficiency. By proactively identifying poor service quality users for the home broadband service and resolving potential issues before user complaints, the solution decreases user complaints more than 25%, reduces on-site processing work orders by 35%, and increases fault processing efficiency by 60%. It also enhances customer satisfaction (NPS) with home broadband services. This project was honored with the "Outstanding FTTH Service" award at the 2024 Network X Conference organized by Informa.

4.5 Wireless Network Energy Efficiency Optimization

Case Background

Energy saving is not only a matter of OPEX reduction for telecom operators, but also endorses that digital technologies play a critical role in achieving carbon neutrality and are integral to addressing climate change globally. Striking the balance between energy-saving functions, while protecting user needs and usage experiences requires intelligent and predictive technology, to minimize the energy consumption per bit at constant user network experience.

Solution

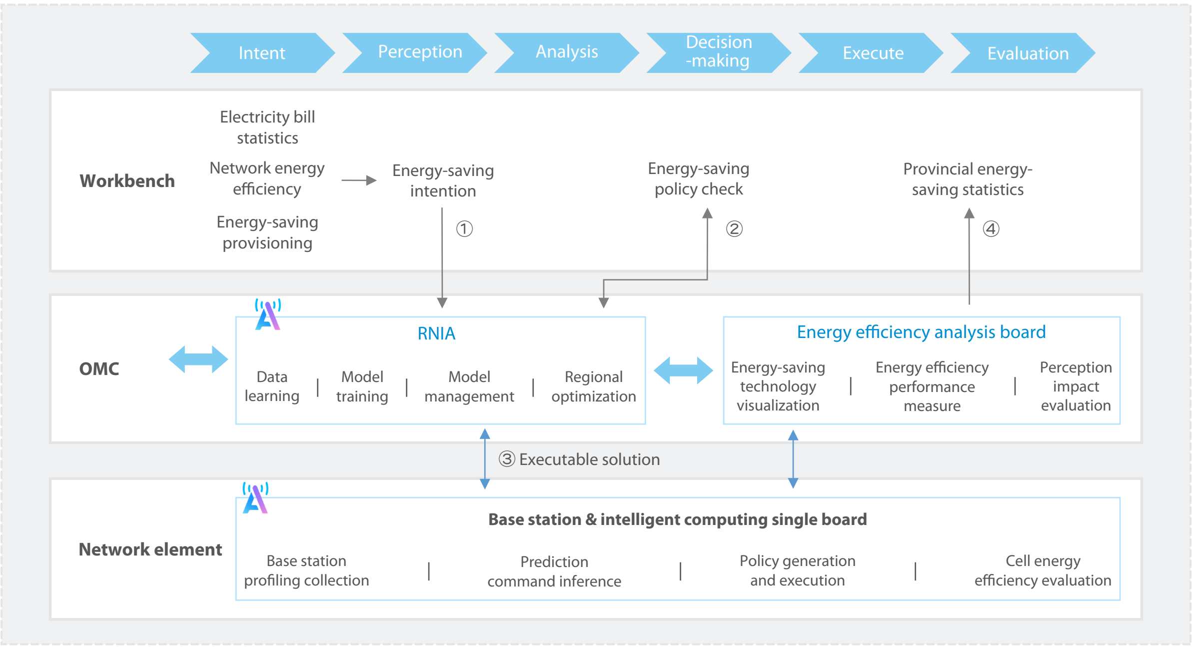


Figure 13. Distributed Architecture of Dual-Layer Endogenous Intelligent Energy-Saving Solution

Energy-saving technologies for wireless network base stations include symbol shutdown in the time domain, carrier shutdown in the frequency domain, channel shutdown in the spatial domain, downlink power control in the power domain, as well as deep sleep modes at the equipment level. However, in practice, wireless networks face objective challenges such as multi-mode, multi-frequency networking, complex service scenarios, and varying user perception requirements. These factors make it difficult to implement energy-saving solutions with precise and efficient "one station, one time, one policy" approaches.

ZTE has pioneered a dual-layer endogenous intelligent energy-saving solution. As shown in the figure, the network management system supports model training, regional optimization, and data visualization, while the base stations process tasks such as profile collection, command inference, and policy execution. Together, they form a distributed architecture design. This solution integrates an endogenous AI-based load prediction model and near-end inference, leveraging innovative technologies such as predictive optimization, autonomous orchestration, multi-frequency coordination, and digital twin visualization. These advancements enable RAN network energy efficiency optimization to achieve "self-learning, self-prediction, self-optimization, self-configuration, and self-evaluation." In collaboration with the operator's workbench, the solution ultimately achieves reduced energy consumption, enhanced energy efficiency and advanced autonomy, without compromising user experience.

Using load prediction technology, future load trends are predicted based on historical data, accurately determining energy-saving periods. Parameter optimization technology automatically identifies the balance between energy savings and network performance, maximizing energy-saving efficiency. Additionally, fast evaluation assurance technology monitors network performance in real-time to ensure that energy-saving measures do not affect user experience. The comprehensive application of these technologies significantly improves energy-saving efficiency while ensuring network performance.

Results

As a concrete example, ZTE has successfully deployed this solution in Liaoning province in China, improving the overall 5G network energy efficiency ratio by approximately 25%. This achievement plays a key role in the evolution of intelligent energy saving in wireless networks and effectively promotes the green and sustainable development of wireless networks.

4.6 SPN Energy Efficiency Optimization

Case Background

Whilst RAN typically represents ~70% of the energy consumption in a telecom operator network, other components, and in particular the transmission network, deserve attention as well and have vast potential for energy savings (annual electricity consumption of 47,157 KW, and electricity costs of around RMB 322 million in China). However, there is a lack of effective energy consumption management methods. Taking SPN as an example, it is mainly reflected in:

- The live network traffic shows a clear tidal effect, with significant differences in load between peak and off-peak times.
- There is a low correlation between service load changes and network equipment power consumption. The equipment often operates at high power consumption levels for extended periods, and a large amount of idle network resources are running on the network, leading toineffective energy consumption.

The question of how to build a smart energy-saving SPN green network, considering the network service features and equipment features, has become a key focus for telecom equipment vendors and operators.

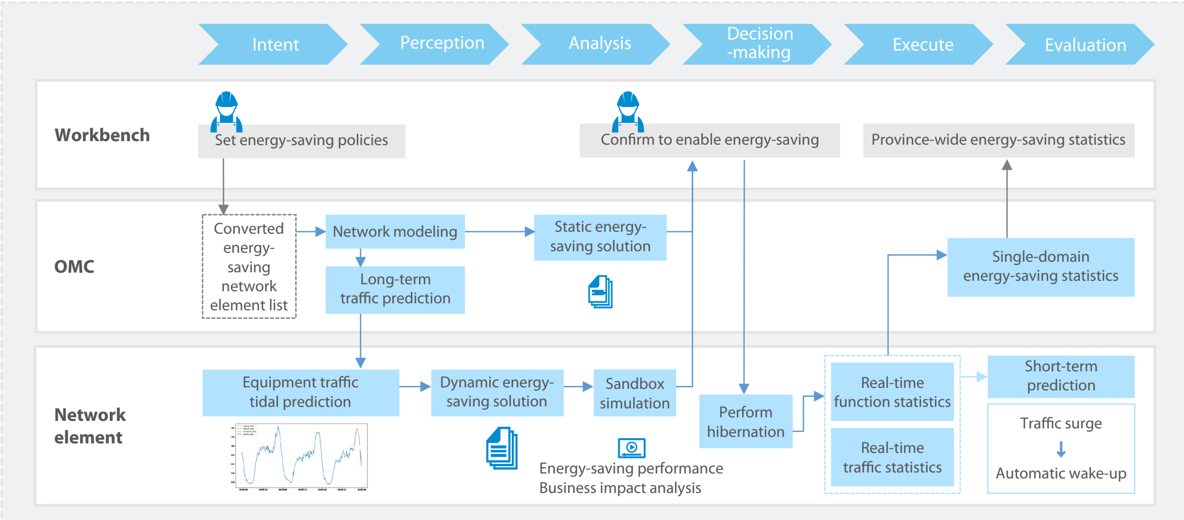


Figure 14. SPN Smart Energy-Saving Solution Diagram

Solution

ZTE, in collaboration with China Mobile, has launched the first SPN systematic smart energy-saving solution. The solution considers multi-dimensional analyses such as equipment type, energy consumption distribution, business traffic analysis and prediction, and algorithm optimization to achieve a three-level energy-saving systematic management at network, equipment, and chips levels. Specific measures include :

- Chip technology: higher integration reduces power consumption per Gbit.
- Partitioned power supply and intelligent power management: ensuring that boards with no services loaded are either not powered or swithed to sleep mode, reduces overall system power consumption.
- Independent air duct cooling and intelligent fan speed control.
- Big data and AI analysis at both control and equipment sides: based on the business traffic and busy/idle conditions of equipment during different periods, intelligently control the dynamic sleep and activation of equipment function modules.

This smart energy-saving solution in the SPN field features multiple innovative technologies, achieving significant energy-saving effects while ensuring business security:

- Built-in AI algorithms realize dual collaboration between network resources and network management. This can be used to analyze or predict the tidal effect of real-time service traffic over time and periodic changes in service traffic patterns (e.g. holidays, major events), and intelligently control the dynamic sleep and startup of equipment chips, modules, boards, and network protection function modules.
- A unique sandbox model simulates energy-saving effects, predicting risks and issues.
- Combining long and short term analyses, advanced service traffic prediction algorithms accurately predict traffic changes. For modules and single boards in sleep mode, they exit sleep before service changes, ensuring service security.
- During the sleep period of equipment chips, modules, and single boards, the system automatically detects the operational

and health status of the components in sleep mode, ensuring the security of the equipment in sleep mode, and allowing them to resume service normally when exiting sleep mode.

- The system displays the energy consumption change of equipment in real-time, and counts the accumulated energy-saving effect of the network on a daily, weekly, and monthly basis.

Results

This smart energy-saving solution has won several prestigious awards, including the Best Application Award and Popularity Award of TM Forum Catalyst Project, SNAI Best Case, Second Prize of "New Green Cup" National Finals, and First Prize in the special competition, demonstrating its leadership position in the industry. Concrete implementations across 29 provinces throughout China resulted to date in an energy-saving impact over 15%, saving nearly 20 million kWh annually.

05

Autonomous Network Evolution Industry Collaboration Initiative

The autonomous network, as a continuously iterating and cyclically evolving system engineering, depends on the prosperity and continuous progress of the ecosystem. ZTE hopes to collaborate with industry chain partners: driven by the end-to-end effectiveness of value scenarios, jointly overcoming core technical challenges and building a solid autonomous network platform engine; adhering to the principle of open decoupling, promoting the innovative development of autonomous services, expanding the boundaries of autonomous capabilities, and jointly creating new value for the communications industry.

Accelerating the development of the autonomous network industry alliance: The autonomous network evolution towards Level 4 (Highly Autonomous) requires the joint efforts from standardization organizations, operators, equipment vendors, scientific research institutions, and the entire industry chain to define a unified industry reference architecture and interface standards, and establish a practical rating system to promote the collaborative development of the industry. As an active practitioner and advocate of autonomous networks, ZTE will actively participate in the discussion and formulation of standards, comprehensively promote highly autonomous network capabilities in value scenarios, assist domestic operators in building autonomous processes, establish international benchmarks, and jointly promote the definition of the autonomous network's target status and industry development.

Fostering innovation and exploration of core technologies: Promote the coordinated development of big data, LLMs, and digital twin technologies, jointly build the autonomous network platform engine, and lay a solid foundation for the intelligent evolution of the network.

- Big data platform: Explore lakehouse and digital intelligence co-governance, providing real-time perception, and multi-modal high-quality corpus supply, achieving AI2Data & Data2AI flexible digital intelligence platform base.
- LLM: Tackle problems such as ultra-long signaling token input, optimization of generalization capability degradation, and multi-modal alignment; improve intent understanding accuracy and cross-domain LPI call accuracy, achieving collaborative intelligent processing of complex network problems.
- Digital twin: Study innovative technologies such as spatiotemporal twin models, real-time service perception twins, and twin agents to create virtual networks, enable virtual-physical mutual control, verify optimal solutions for network O&M iteration, and feed it back into the real network to form a closed-loop optimization.

Accelerating the deployment and application of Level 4 autonomous networks: Use value scenario applications, explore new algorithms and applications, gradually pilot them, and expand application scenarios and scale. First, achieve full autonomy in single-domain networks, and gradually evolve to cross-domain collaboration. Accelerate autonomy deployment, improve network quality, increase operational efficiency and resource utilization, ensure excellent user experience, and expand the network's value.

ZTE will cooperate with operators and relevant vendors to carry out the following work to continuously promote the implementation of Level 4 autonomous network:

- Sorting out the end-to-end O&M operation process.

- Combining with existing operator systems, eliminating process bottlenecks and gaps, integrating into the production process, and improving automation capabilities.
- Refining effectiveness metrics, ensuring the achievement of value scenarios, and implementing and promoting them.
- Lowering the skill threshold for network operations, gradually replacing manual work, and realizing autonomous closed-loop of the network through key intelligent technologies such as the collaboration of large and small models, digital twins, and agents.

Looking ahead, ZTE is willing to communicate and cooperate closely with operators and partners, work together, promote the vision and concept of autonomous networks to become an industry consensus, continuously explore and practice technologies, accelerate the value transformation of autonomous networks, and empower the digital economy and industry digital transformation.

Acronyms

Acronyms	English
5G-A	5G-Advanced
GenAI	Generative AI
AN	Autonomous Networks
ANL	Autonomous Networks Level
API	Application Programming Interface
CCSA	China Communications Standards Association
CHB	Customer, Home, Business
DTW	Digital Transformation World
ENI	Experiential Networked Intelligence
ETSI	European Telecommunications Standards Institute
IDMS	Intention-driven management System
ISG	Industry Specification Group
ITU	International Telecommunication Union
KBI	Key Business Indicator
KCI	Key Capability Indicator
KEI	Key Effectiveness Indicator
LLM	Large Language Model
MTL	Metrics， Tracing， Logging
NGMN	Next Generation Mobile Network
NOC	Network Operation Center
ODN	Optical Distribution Network
OMC	Operation and Maintenance Center

Acronyms	English
O&M	Operations and Maintenance
OPEX	Operating Expense
OTN	Optical Transport Network
PON	Passive Optical Network
RAG	Retrieval Augmented Generation
SOC	Service Operation Center
SPN	Slicing Packet Network

References

[1] China Academy of Information and Communications Technology. Global Digital Economy White Paper (2024)

[2] China Academy of Information and Communications Technology. Research Report on the Development of China's Digital Economy

[3] TM Forum. Autonomous Networks: Empowering digital transformation – evolving from Level 2/3 towards Level 4

[4] TM Forum. TMForum Autonomous networks：Level 4 industry blueprint

[5] TM Forum. TMForum__Autonomous_networks_level_4_industry_blueprint_11_24

[6] China Mobile. China Mobile Autonomous Network Whitepaper (2024)

[7] China Telecom. China Telecom Cloud Network Operation Autonomy White Paper 4.0 (2024)

[8] China Unicom. China Unicom Autonomous Network White Paper (2024)

[9] Gartner. Hype Cycle for Generative AI, 2024

[10] OpenAI. https://openai.com/index/introducing-openai-o1-preview/

ZTE CORPORATION

NO.55, Hi-tech Road South, ShenZhen, P.R. China
Postcode: 518057
Tel: +86-755-26770000
Fax: +86-755-26771999
Website: www.zte.com.cn