ZTE

ZTE Autonomous Networks White Paper 2026

Ecosystem co-building Intelligent Connection Future

Reflections and Practice of L4 High-Level Autonomous Networks

www.zte.com.cn

Ecosystem co-building
Intelligent Connection Future

Reflections and Practice of
L4 High-Level Autonomous Networks

CONTENT

Abstract

Thanks to the large-scale commercialization of 5G-A, the deep integration of AI LLMs into communication systems, and the maturing of digital twin technology, global operators are now able to realize the value of L4 high-level autonomous networks for tackling key business issues. Autonomous networks are no longer limited to operation and maintenance automation, but are transitioning towards an intelligent closed-loop system of real-time perception, active analysis, autonomous decision-making, and automated execution. They are gradually becoming the core engine for operators to achieve service revenue growth, service innovation, and green low-carbon development. TM Forum has clarified that 2025-2027 is a critical stage for the autonomous closed-loop of L4 single-domain maintenance, optimization, and operation scenarios. Standard organizations such as CCSA are accelerating the promotion of L4 core technical standards, including LLMs and agents, further consolidating the foundation for industrial evolution.

ZTE has been deeply involved in the autonomous network industry for many years. Based on the accumulation of L3 large-scale applications, it actively promotes L4 highly autonomous R&D and pilots. Through full-stack AI architecture evolution and value scenario expansion, it promotes the paradigm upgrade of the network from improving operation and maintenance efficiency to achieving revenue growth. Based on ZTE's L4 pilot practices with global operators, this White Paper analyzes the core capability requirements and industrial evolution path of high-level autonomous networks, and describes ZTE's L4 high-level evolution solution, characterized by value-driven, endogenous intelligence, autonomous decision-making, and automated closed-loop.

To present the evolution path and implementation plan of L4 high-level autonomy, the White Paper will focus on the following four aspects:

1. Insights into the evolution of L4 high-level autonomous networks:

Based on security and controllability, this section explores the continuous evolution of the full-stack AI architecture by applying advanced technologies such as LLMs, agents, and digital twins. It outlines the construction of work order-free end-to-end operation and maintenance capabilities featuring real-time perception, active analysis, autonomous decision-making, and automated closed-loop, and expands into high-value scenarios of market-network-service collaboration, implementing a value-effectiveness closed-loop from traditional Key Performance Indicator (KPI) to Key Business Indicator (KBI), Key Effectiveness Indicator (KEI), and Key Capability Indicator (KCI).

2. ZTE L4 high-level autonomous network evolution solution:

Focusing on the AIR Net full-stack AI architecture upgrade, this section details how to achieve autonomous decision-making and automated closed loops, upgrading from operation and maintenance efficiency enhancement to value-based operations to empower operator revenue growth and ecosystem win-win. The solution comprises six core dimensions: (i) placing value orientation at the core to deepen the KBI-KEI-KCI effectiveness system for driving quality improvement, efficiency enhancement, cost reduction, and revenue growth; (ii) advancing the Agentic AI architecture to realize layered intelligence injection and end-to-end collaboration; (iii) upgrading intelligent hardware with native network element (NE) intelligence to solidify NE-level real-time sensing and remote control capabilities; (iv) evolving the ZTE Nebula Telco Large Model into a "1+N" model matrix featuring key AI innovations such as Knowledge Graphs, Graph Search, and the Co-Sight Agent Factory, which, combined with digital twins, upgrades AI from auxiliary support to autonomous decision-making; (v) reshaping processes to cut through breakpoints and bottlenecks for advancing end-to-end work order-free operation and maintenance; and (vi) expanding high-value market-network-service collaboration scenarios to enable business revenue growth.

3. ZTE autonomous network practices:

This section highlights typical L4 practice cases, covering scenarios such as network fault handling, network changes, network optimization, complaint handling, and hierarchical value monetization. These cases illustrate the dual value of high-level autonomous networks in terms of business effectiveness and operational efficiency.

4. Industrial collaboration initiative:

Moving forward, it is advocated that all parties in the industry chain go hand in hand, taking standard co-construction, technology co-research, scenario co-creation, and ecological win-win as the core paths, accelerating the maturity of L4 autonomous network standards, building a strong AI security governance framework, exploring new network business models, helping global communication networks enter a new stage of L4 high-level autonomy, and thus enabling the whole industry to share the dividends of high-level autonomous networks development.

01

Insights into the Evolution of L4 High-Level Autonomous Networks

With the surge in communication networks complexity and the continuous expansion of service scenarios, traditional operation and maintenance models have struggled to meet efficient, agile, and intelligent operational needs. The evolution of autonomous networks to L4 high-level is not only a necessary path for network technology upgrades, but also a key leap for operators to move from cost reduction and efficiency improvement to intelligent operation. In 2025, driven jointly by standard organizations, operators, and suppliers, L4 high-level autonomous networks have made significant progress in standards, systems, and practices. At the same time, higher requirements are placed on L4 high-level autonomous networks. In this context, gaining deep insights into the development trends, key technologies, and security challenges of L4 high-level autonomous networks is of great significance for grasping the future direction of network intelligence and exploring new growth opportunities.

1.1 L4 High-Level Autonomous Networks Rely on Technologies such as LLMs, Agents, and Digital Twins

Insight and Analysis

The main feature of L4 autonomous networks is the organic closure of end-to-end loops based on intelligent hardware real-time perception, precise analysis, intelligent decision-making, and online execution. Achieving this goal requires the collaborative empowerment of three key technologies: LLMs, agents, and digital twins. LLMs serve as the core cognitive reasoning hub, with multi-agents as collaborative closed-loop execution units, and digital twins as decision verification sandboxes. The three form a complete technical ecosystem, helping operators transform from auxiliary automation of script-dependent, machine-assisting-human to autonomous decision-making of intent-driven, human-supervising-machine.

Cognitive upgrade:

The LLM acts as the network "super brain", able to understand vague natural language intents (such as "improving user experience in XX area"), accumulating expert experience to handle root cause analysis and optimization of complex networks, and extracting deep semantic associations from massive, multi-modal operation and maintenance data. This capability upgrades operations from executing instructions to a paradigm of understanding intent, autonomous reasoning, and long-term planning, which is the foundation for achieving cross-domain and cross-scenario high-level decision-making.

Execution reconstruction:

The agent is no longer a simple Copilot (auxiliary tool), but an intelligent unit with autonomous decision-making, automated closed-loop, and self-learning evolution capabilities. Furthermore, multi-agent collaboration helps realize cross-domain and cross-scenario collaboration, such as cross-domain end-to-end closed-loop from user to equipment. The end-to-end management of network operation and maintenance services involves multiple technologies such as wireless, core network, and transmission. Siloed single-domain, single-scenario intelligence cannot solve cross-domain collaborative optimization problems. Multi-agent clusters are the key execution architecture for achieving cross-domain self-healing. Through a unified telecom-grade multi-agent communication protocol, agents in various domains and single scenarios can break boundaries, negotiate dynamically, share status, and jointly complete task decomposition and execution under complex multi-objective constraints, achieving true cross-domain end-to-end and cross-layer (business, service, equipment) end-to-end dual closed loops.

Risk pre-positioning:

Digital twins build the security foundation for autonomous decision-making. The zero-touch requirement of autonomous networks demands high system reliability, while the black box risk of AI decision-making and the potential impact on live network operations are currently the biggest obstacles hindering the implementation of closed loops. Digital twins establish precise network images, which are the critical basis for verifying network decisions. Network element endogenous twins reflect equipment status in real-time, and service layer twins simulate user behavior and network response. By seeking global optimal solutions across dimensions such as capacity, latency, and energy efficiency through multi-objective collaborative optimization, predictive decision-making capabilities support early warning and active planning of capacity bottlenecks.

Recommended Solutions

To accelerate the transition towards full-scenario L4 high-level autonomy, operators should focus on the following capability construction and architecture transformation:

Integration transformation based on high-value scenarios: Reshape operation and maintenance processes to achieve a shift from passive response to active foresight. At the same time, prioritize upgrading cross-domain and cross-scenario high-level autonomous capabilities in high-value, high-frequency scenarios such as network fault handling, change monitoring, and service quality optimization. The key to process reshaping based on high-value scenarios lies in the deep integration of LLM reasoning, digital twin risk verification, and multi-agent collaborative closed-loop, ensuring that all high-level policies are rigorously rehearsed in a sandbox before execution and can complete dynamic execution closed loops.

Promote standardization and security of agent systems: Accelerate the promotion of cross-domain agent collaboration mechanisms of telecom-grade multi-agent protocols, promote the standardization of protocols in standard organizations such as TM Forum and CCSA, and break down vendor barriers. At the same time, strict agent operation security specifications and control mechanisms, such as electronic fences, must be established to ensure that the autonomous evolution of agents operates within preset behavioral boundaries and is monitored and guided by human experts, effectively controlling the potential risks of AI execution.

Promote layered and graded deployment policies: Based on the deployment location of the LLM (cloud-edge-end), build model lightweight policies to gradually promote the reduction of reasoning costs. Implement lightweight digital twin policies, adopt multi-dimensional precision characterization, and apply differentiated model details and computing resources according to the importance of different regions or objects. For example, high-fidelity modeling is performed for core domains (such as internal equipment of 5G base stations, antenna arrays), while highly simplified abstract models are adopted for edge domains (such as large signal coverage areas).

1.2 L4 High-Level Autonomous Network Systems Will Evolve to a Full-Stack AI Architecture

Insight and Analysis

AN L4 capability demands drive architecture from local AI to full-stack AI

L4 autonomous networks pursue the core goals of end-to-end autonomous collaboration and intent-driven. Traditional local AI deployment has obvious shortcomings. Low-level autonomous networks' AI is mostly concentrated in operation and maintenance, and the network element layer lacks endogenous intelligence, leading to lagging data perception and poor cross-domain collaboration, unable to support full-link capabilities of real-time perception, autonomous decision-making, and automated closed-loop. The full-stack AI architecture achieves comprehensive intelligent capability penetration through the vertical integration of network elements, platforms, and applications. Network elements have autonomous perception and execution capabilities relying on small-sized network element models for reasoning, the platform relies on communication LLMs for overall decision-making, and applications accurately map service intents, bridging the capability gap from single-domain intelligence to cross-domain collaboration and meeting the core requirements of the L4 stage defined by TM Forum.

The AI technology graph capability is gradually maturing, providing feasibility support for the full-stack architecture

Technological breakthroughs represented by generative AI and Agentic AI have moved full-stack AI deployment from theory to reality. Communication LLMs with hundreds of billions of parameters solve the problems of network knowledge understanding and policy generation, realizing natural language interaction and adaptation to complex scenarios; multi-agent collaboration technology supports inter-domain distributed decision-making, and Chain of Thought (CoT) capabilities enhance the logic of autonomous decision-making; the combination of AI acceleration chips and endogenous intelligent network elements reduces edge AI inference latency and ensures real-time response; network element small-size models provide inference model capabilities in embedded base station systems or resource-constrained scenarios; the Ray distributed framework provides distributed scheduling and computing management of infrastructure for unified management of heterogeneous computing power such as CPU/GPU. These AI technology graph capabilities are gradually maturing and working collaboratively to break the limitations of fragmented algorithms and weak scenario adaptation of traditional AI, building the foundation for full-stack AI technology.

Industrial practice verifies value, promoting consensus towards full-stack AI architecture

Operators' practices have fully verified the feasibility and value of full-stack AI. TM Forum proposed an Agent-based Agentic AI architecture, whose agents double the efficiency in scenarios such as fault localization and energy saving optimization; ZTE launched the first autonomous agent and Intelligent-Digital Employee solution based on LLMs and agents. These practices prove that the full-stack AI architecture can not only reduce operation and maintenance costs but also activate the potential of network revenue growth, becoming the core choice for operators to move towards the L4 stage. At the same time, TM Forum released the AN L4 industry blueprint to further promote full-stack AI as the unified evolution direction of the industry and reduce resistance to cross-vendor collaboration.

In summary, the full-stack AI architecture not only breaks through the core technical bottlenecks of the L4 stage, but also matches the business needs of operators to reduce costs and increase revenue. This has become the industry-recognized evolution goal of L4 autonomous networks, promoting the network leap from machine-assisting-human to human collaborating with machine, and shifting the network operation and maintenance paradigm from the rule of man to agent collaborative autonomy.

Recommended Solutions

Operators should build a layered collaborative architecture centered on full-stack AI and combined with the evolution of network agentification. By leveraging network atomic capabilities, network AI, operation and maintenance AI, digital intelligence engines, and the future intelligent hub, the system supports network agent and multi-agent collaborative architectures to achieve automation and intelligence. This evolution follows the path from network element to network management, from network to service, and ultimately to future full-stack AI.

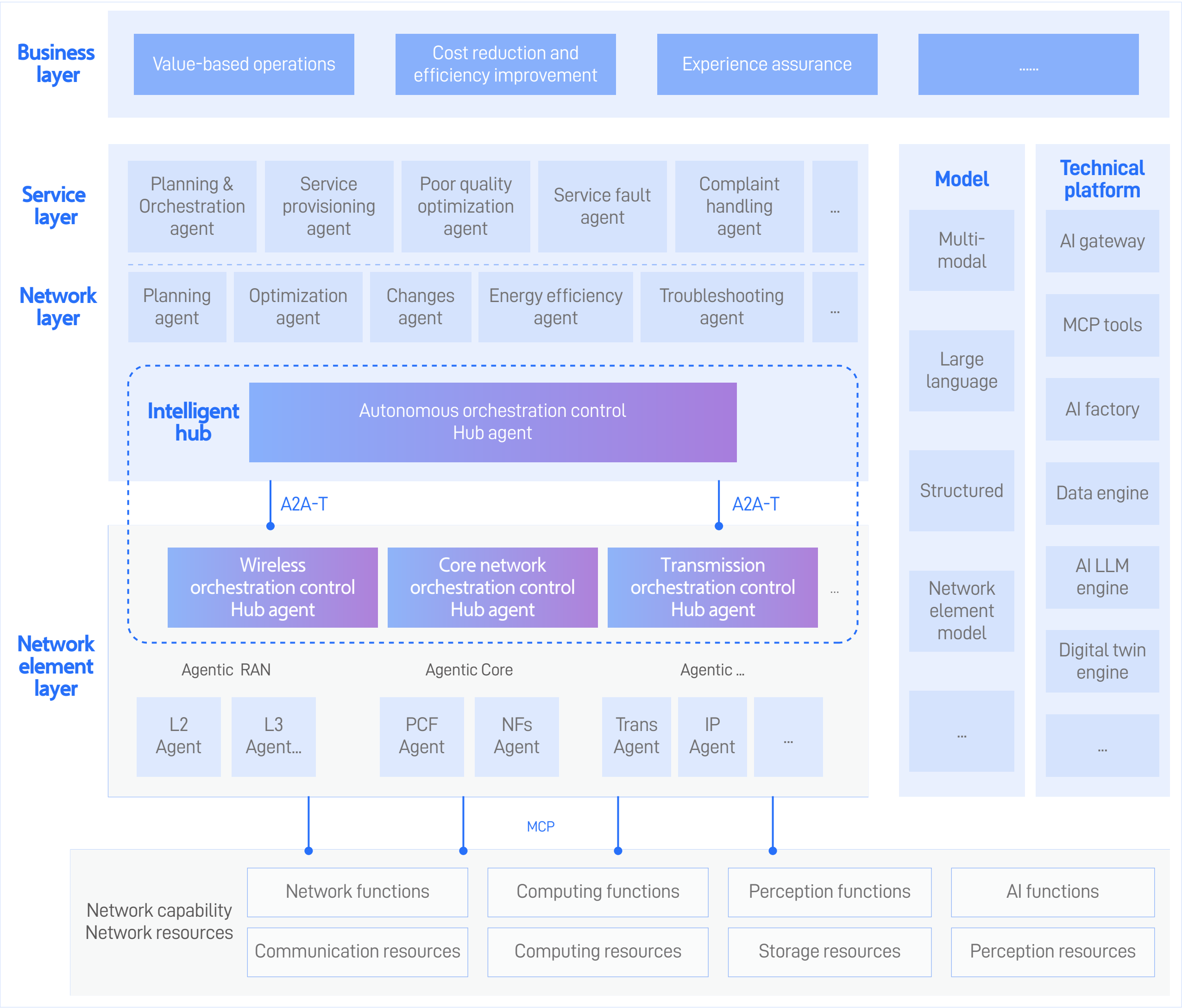

Figure 1-1 Insights into Full-stack AI Architecture Evolution

The full-stack AI architecture is not just about using a few AI algorithms in the network. It is a comprehensive system that runs through the whole process with layer-to-layer agent collaboration. It can be generally divided into the network element layer, network layer, service layer, and business layer.

Network element layer: corresponds to the agentification of network element system equipment, such as wireless, core network, transmission & IP network, etc. Through small-size LLMs and network models with exclusive model reasoning capabilities, collaborating with a series of built-in agent applications in the network, operators deploy intelligent hardware in the network or build AI chips and models into network equipment (such as base stations, core networks, routers) to process local perception data in real-time and make quick decisions, such as wireless resource allocation, traffic scheduling, beam prediction, routing forwarding, etc. In the future AI-based network, network capabilities and network resource capabilities will use MCP, to provide rich atomic capability services, including network functions, perception functions, AI functions, etc., as well as various communication, storage, and computing resources, forming a network MCP-enabled atomic capability and tool ecosystem.

Network layer: Achieve agentification of network management through agent online memory and workflow orchestration, single-agent reasoning, tool calling, and multi-agent collaborative work capabilities. LLMs provide core capabilities for high intent recognition and Text-to-SQL processing accuracy, support signaling analysis, adopt key technologies such as multi-LoRA acceleration and context compression, provide network layer agent applications, and support network operation and maintenance agent applications and collaborative work including network planning, network optimization, network change, and energy efficiency optimization. The intelligent hub is a vision for future evolution, a ubiquitous concept. In fields such as wireless, core network, wired, and autonomous networks, a "super-agent" responsible for the orchestration and control hub may develop in the future. This hub agent is responsible for the collaboration and orchestration of various agents in its domain and collaborates with other hub agents. As the intelligent hub after the future network full-stack AI-fication, it realizes network operation, resource scheduling, service orchestration, etc., acting as the network brain.

Service layer: Relying on agents and multi-agent collaboration for planning and orchestration, service provisioning, poor quality optimization, service failure, complaint handling, etc., operators should build an intelligent service system, forge end-to-end full-chain service support capabilities, realize service agentification, cover customer needs, and achieve a "demand / service" closed-loop; at the same time, operators should promote efficient service response, compress the customer demand response cycle to the maximum extent through measures such as core capability forward movement, time-consuming stage optimization, and process simplification, laying a solid intelligent foundation for high-quality service operation. In the future, with the deep integration of network management and service operation, an integrated intelligent hub for network operations and maintenance will gradually take shape.

Business layer: The iterative upgrade of LLM and agent technologies will also drive the agentification of value-based operations, cost reduction, efficiency improvement, and user experience assurance, injecting new intelligent momentum into commercial operations. On the one hand, it can explore new growth points of commercial value and broaden the path of increasing business revenue; on the other hand, it optimizes the cost structure, consolidates service quality, improves user satisfaction, and matures the commercial realization of user experience.

Technical platform: Operators should build a unified technical platform, including AI gateway, MCP tool ecosystem, AI factory framework, etc., to support calls between AI applications and LLMs, between AI applications and tools, and between AI applications and AI applications, organically integrating models, tools, and data flows, understanding and executing dynamic tasks generated by models, and providing stable, efficient, and secure guarantees for the entire process. The base of the technical platform is the digital intelligence engine composed of three major technologies: data engine, AI LLM engine, and digital twin engine. The three collaborate and support each other. Among them, the data engine is the foundation, providing required data resources for the AI LLM engine and digital twin engine, and ensuring effective data governance. The AI LLM engine is the intelligent support for the future network, supporting the construction of digital twin models by training scenario LLMs. At the same time, the digital twin engine can generate new data, which can, in turn, further iterate training and optimize the LLM. This mutually empowering relationship ensures that the three technological engines work closely together, driving technological innovation and development.

1.3 L4 High-Level Autonomous Networks Need Key Capabilities Such as Real-Time Awareness, Active Analysis, AutonomousDecision, and Automated Closed-Loop Execution

Insight and Analysis

L4 high-level autonomous network is a major leap for network autonomy towards the stage of true autonomous decision, capable of autonomously completing tasks and workflows in a closed loop according to user intent without human intervention. In 2026, the evolution of L4 high-level autonomous networks will enter its most challenging phase, further pursuing real-time network awareness, active analysis, autonomous decision, automated and intelligent closed-loop, and deeply integrating cutting-edge technologies such as LLMs, agents, and digital twins to build a full-stack AI-driven autonomous closed-loop architecture.

According to TM Forum's definition, the autonomous closed-loop framework includes the following stages: Goal and Intent Management (Intent), Identification (Awareness), Scoping and Positioning (Analysis), Solution Generation and Decision, Handling and Verification (Execution).

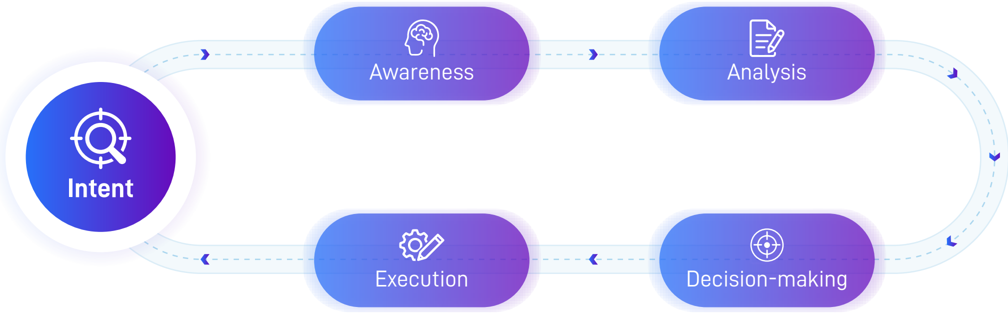

Figure 1-2 AN Autonomous Closed-Loop Framework

The awareness stage actively captures dynamic information such as stage status, service traffic, and equipment payload in real-time to build a solid data foundation for subsequent intelligent operations. The analysis stage deeply mines awareness data through AI, which can not only accurately identify known problems, such as network congestion and hidden failure dangers, but also predict potential risks, such as service peaks and resource bottlenecks, realizing an upgrade from passive response to active prediction. The decision stage is the core of the autonomous closed-loop. Without human intervention, resource allocation is dynamically optimized, network parameters are adjusted, and redundant stages are switched based on analysis results to ensure the accuracy and timeliness of decisions.

L4 high-level autonomous networks will leverage Agentic AI as the core enabling technology for their autonomous closed-loop framework. The agent can perceive environmental changes promptly, analyze and make decisions based on its learned knowledge, and take action to achieve specific goals, possessing autonomy, adaptability, and intent management capabilities. Therefore, Agentic AI is highly matched with the goals of the L4 high-level autonomous network closed-loop framework.

Recommended Solutions

Real-time Awareness based on intelligent hardware: Existing communication equipment often only focuses on the implementation of communication protocols in design, without much consideration for data collection. Real-time collection of large amounts of data may affect the normal operation of services. To ensure real-time and accurate awareness without affecting services, operators should consider introducing new intelligent network element equipment, such as intelligent computing boards, to actively collect large amounts of information, such as user and equipment internal status in real-time, thereby quickly identifying various faults and anomalies.

Context management based on event aggregation: In the awareness stage, to provide more real-time and accurate identification capabilities, it is necessary to quickly collect various data from multiple data sources, filter and correlate data, aggregate multiple anomalies into events, provide the most appropriate information input for subsequent automated analysis and processing, and avoid information omission or overload.

Joint analysis based on multi-agent collaboration: In the analysis stage, the analysis of problems is quickly completed through the collaboration of cross-domain scoping agents and single-domain positioning agents. Technologies such as knowledge graphs can be used to systematically construct and manage knowledge across individual domains. This approach ensures consistency and completeness in both cross-domain and single-domain analyses, while also preventing the fragmentation of domain-specific knowledge.

Assisted decision based on digital twin and knowledge base: In the decision-making stage, operators should strengthen the accumulation of the knowledge base and case experience so that AI can generate more reasonable solutions. For solutions, the effects can be simulated through technologies such as digital twins to improve the credibility of decisions.

Data-driven autonomous iteration: In the execution stage, it is necessary to consider minimizing the impact of network operations on services. Operators should try to avoid global one-time operations, shorten service recovery time during faults, isolate faults, and dynamically adjust execution plans or update solutions based on real-time feedback of network conditions.

1.4 L4 High-Level Autonomous Networks Are No Longer Limited to Operational Efficiency, but Will Explore New Scenarios and Opportunities for Business Revenue Growth

Insight and Analysis

Facing the dilemma of fading traffic dividends and the scissors difference of increase in volume but not in revenue, the traditional operation cost reduction and efficiency improvement initiatives of global telecom operators have reached the limits of marginal benefits. Mere improvement in operational efficiency is no longer enough to hedge against the pressure of stagnant revenue growth. The core demand of autonomous networks has shifted from how to save money to how to make money.

The high-level evolution and maturity of autonomous networks come at the right time. It is not only an iteration of O&M tools, but also the cornerstone of business model reconstruction. With technological breakthroughs such as intent-driven, real-time perception, and cross-domain closed-loop, the L4 network transforms the invisible black box into perceptible, customizable, and affordable flexible assets, realizing a perfect interlock between technical capabilities and business demands. Empowered by communication LLMs and digital twins, operators will lock in experience dividends and complete the strategic transformation from connection pipe providers to digital service revenue generators.

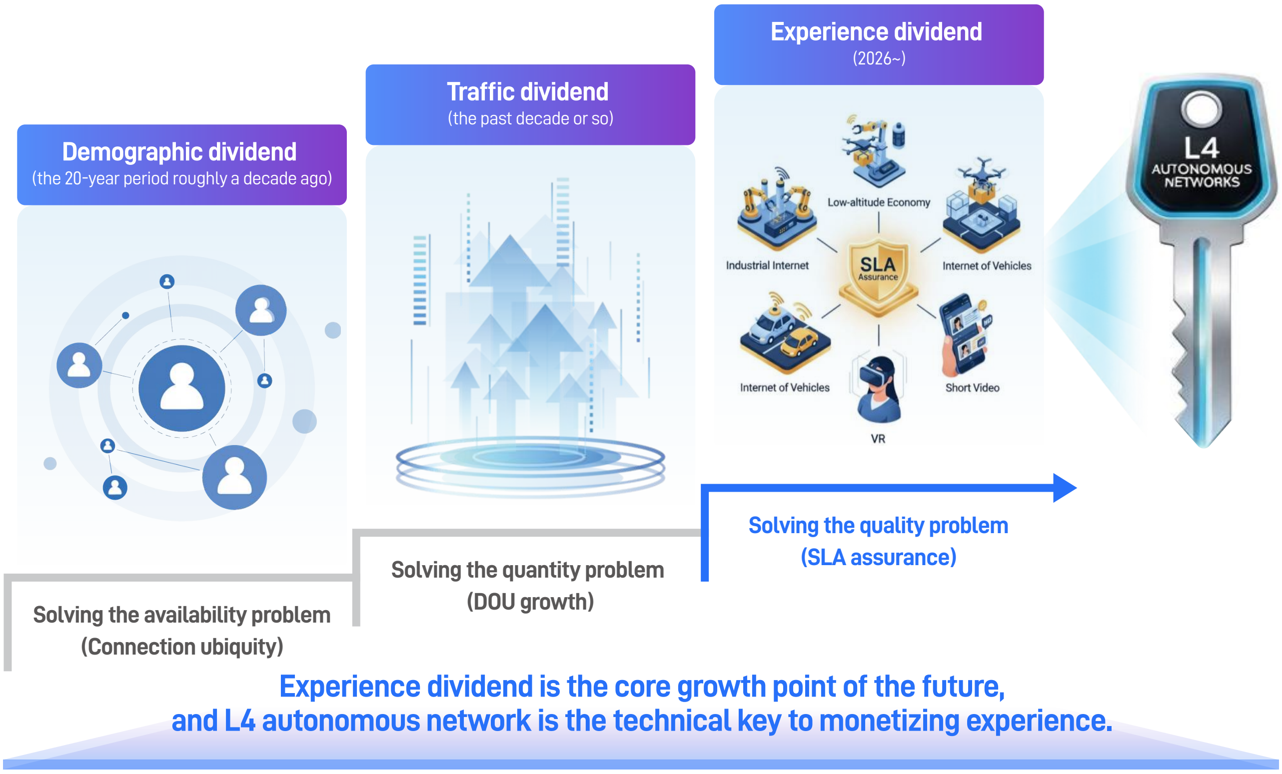

Figure 1-3 Evolution of Dividends in the Telecom Industry

Recommended Solutions

High-level autonomous networks can expand three major revenue growth paths for operators: differentiated experience monetization, deep cultivation of vertical industries, and data element valorization and ecological reconstruction, bridging the vacuum period of traffic dividends, and extending autonomous network construction from the operational dimension of cost reduction and efficiency improvement to the business dimension of revenue growth empowerment.

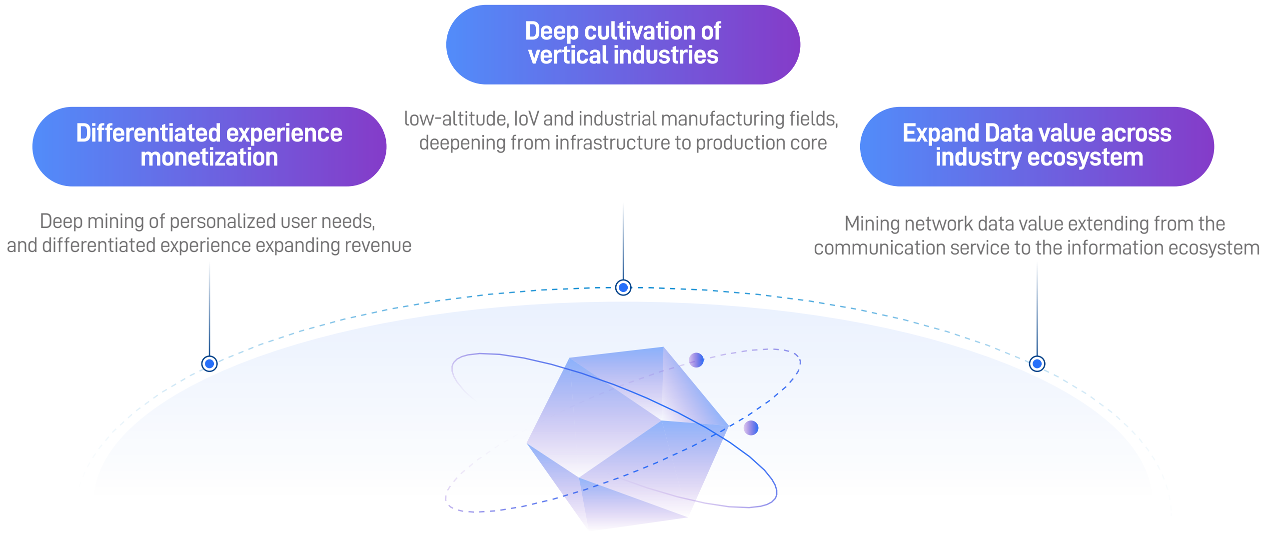

Figure 1-4 Exploration of Revenue Growth Paths for Autonomous Networks

Scenario 01

Differentiated experience monetization:

Moving from Best-Effort to Experience Contract. In the B2C market, operators can rely on autonomous networks to identify high-value users and scenarios, providing deterministic experience assurance.

Cloud gaming/XR: Launch an "Immersive Acceleration Package" to establish dedicated low-latency channels, resolve stuttering and motion sickness, and achieve service premium.

Live streaming: Operators can create Uplink Guarantee Cards, provide dynamic bandwidth and anti-jitter guarantees, and convert ordinary traffic into high-ARPU HD live streaming services.

Scenario 02

Deep cultivation of vertical industries:

Promotion from Infrastructure Provider to Production Partner. In B2B field, operators can provide customized slicing and operation and maintenance stewardship services to adapt to emerging scenarios.

Low-altitude economy: Operators can utilize 5G-A communication-sensing integration technology to provide communication + sensing + navigation services and explore new charging models based on routes/flights.

Industrial/Internet of vehicles: Operators can guarantee the highly reliable operation of smart mines and autonomous driving, launch visual private line stewardship services, and achieve a leap from connection to comprehensive services.

Scenario 03

Data monetization:

Expanding from Communication Services to Information Services.

Data insight: Operators can convert desensitized data into business insights, serving for instance car company marketing, retail site selection, and financial risk control, realizing the monetization of data elements.

Capability exposure: Operators can encapsulate data, service status, QoS, services, and other capabilities into APIs, build NaaS product capabilities, and achieve large-scale capability invocation.

From improving operation and maintenance efficiency to increasing business revenue is the inevitable mission of the development of L4 high-level autonomous networks. Through the precise fit of technological evolution and market demand, operators will completely break the barriers between technology and business. Through differentiated experience, productivity empowerment, and data ecosystem monetization, operators will build a sustainable second growth curve.

1.5 L4 High-Level Autonomous Networks Must Prioritize AI Application Security

Insight and Analysis

For L4 high-level autonomous networks, AI is no longer an auxiliary tool but the core engine of perception, analysis, decision, and execution. The security status of AI directly determines the network's survivability. AI decision deviation or attack may cause catastrophic consequences, such as service interruption.

At the compliance level, domestic and international regulations form a dual constraint: Countries are accelerating the construction of corresponding laws, regulations, and governance systems. China's *Interim Measures for the Management of Generative AI Services and Data Security Law* clearly require that AI applications meet compliance, explainability, and accountability. Massive user communication data and network topology information processed by operators need compliant processing in AI training/inference. Violations can be fined up to RMB 50 million. The European Union's *AI Act* lists communication network operators' AI systems as high-risk, requiring full life-cycle compliance transformation, including data quality verification, algorithm explainability, and human supervision node setting before 2026. Violations can be fined up to 7% of global annual turnover.

The autonomous decision-making capability of agents puts the traditional cybersecurity defense system under comprehensive innovation pressure in terms of defense concepts (from passive interception to active immunity), technical tools (from rule matching to intelligent perception), and security architecture (from single-point protection to collaborative defense).

As large-scale applicants of agent technology and guardians of critical infrastructure, operators' AI security focus is highly focused on actual service pain points: compliance of large-scale data processing, security of high-privilege operations, and core service stability, requiring a balance between AI efficiency and security.

Recommended Solutions

Operators should treat security and compliance as top priority, embed compliance and security requirements into the full AI life cycle, and build an AI security ecosystem in four dimensions:

01 Intelligent threat detection and identification: to accurately intercept traditional and new attacks

AI technologies such as deep learning and LLMs can be used to conduct real-time, in-depth analysis of traffic, behavior, and AI-generated content (such as deepfake audio and video) in communication networks. It can accurately and quickly identify and intercept new attacks, such as prompt injection and deepfakes, while upgrading traditional threat defense to accurately identify telecom network fraud, spam, and malicious content, improving the fraud information interception rate while reducing the false interception rate.

02 Privacy enhancement technology: enabling data availability without data exposure

Operators should adopt privacy enhancement technologies such as federated learning, differential privacy, and secure multi-party computation throughout the full life cycle of AI models and applications. In the training phase, federated learning is adopted to support non-exit and co-training of cross-province/cross-border data; in the inference phase, differential privacy technology is applied to avoid raw data leakage; dynamic desensitization and encrypted storage are implemented throughout the process to meet domestic and international data compliance requirements.

03 Full-cycle AI security assessment mechanism: building a strong AI robustness defense line

Operators should incorporate security assessment throughout the AI lifecycle ("development - operation - retirement"). Specifically, they should embed the Security Development Lifecycle (SDL) process in the development phase and conduct security reviews; simulate attack scenarios through digital twins in the operation phase to verify AI decision security, and regularly conduct adversarial attack and defense tests to improve the anti-attack capability of the model and application itself; and securely destroy models and data in the retirement phase, cancel machine identity permissions, and eliminate legacy risks.

04 Dynamic intelligent collaborative defense: building a cybersecurity brain

Operators should integrate network-wide security data and posture, use AI technology to analyze and judge threat posture, and automatically optimize and adjust security policies. They should realize cross-domain collaborative defense of automated threat discovery and intelligent linkage disposal, upgrade protection from passive response to active defense, and truly reflect the self-healing capability of L4 autonomous networks under human supervision.

02

ZTE L4 High-Level Autonomous Network Evolution Solution

2.1 ZTE L4 High-Level Autonomous Network Evolution

2.1.1 ZTE Autonomous Network Three-Stage Evolution

ZTE uses digital intelligence technology as an engine to empower operators' autonomous network construction. After years of technological innovation, it has successfully transformed autonomous networks from a conceptual blueprint into an indispensable core capability in operators' production and operation. The automation and intelligence solutions provided by ZTE have reached an industry-leading level, laying a solid and reliable technical foundation for operators to move towards high-quality and high-efficiency intelligent network operations.

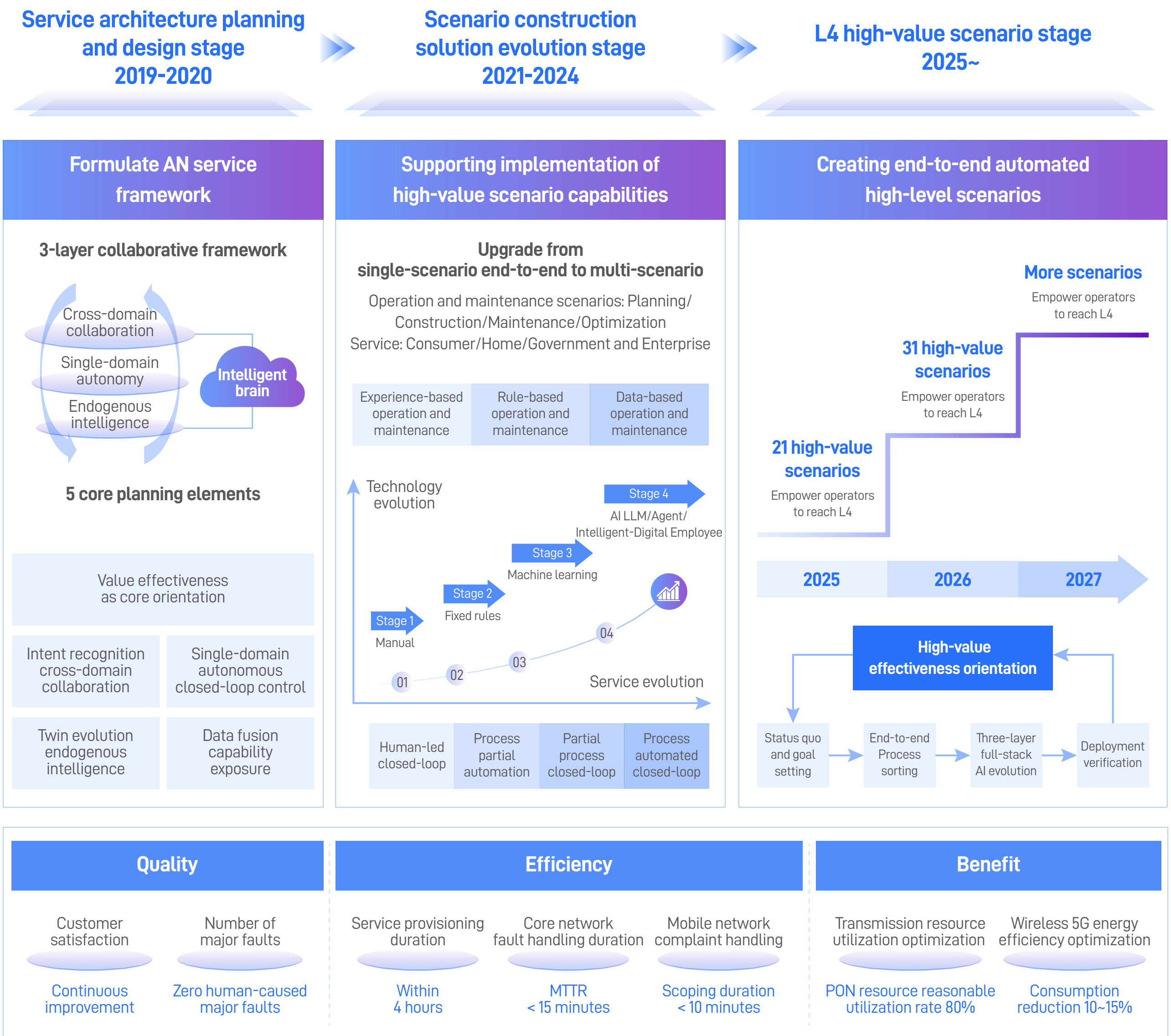

Figure 2-1 ZTE Autonomous Network Evolution

Service architecture planning and design stage (2019-2020): This stage focused on laying the foundation and establishing rules to build the ZTE autonomous network skeleton. It centered on five planning elements, was driven by value effectiveness, and emphasized network element endogenous intelligence, single-domain autonomy, cross-domain collaboration, and data fusion capability exposure. ZTE built a three-layer collaborative framework and formed an organically unified network system to ensure that the autonomous network construction direction is correct, the path is clear, and the capabilities are scalable.

Scenario construction solution evolution stage (2021-2024): Serving as a critical period for moving from blueprint to practice, this stage saw the continuous iterative construction of high-value scenario capabilities and the continuous improvement of scenario construction solutions throughout the full life cycle of planning, construction, maintenance, and optimization, as well as in CHB (Consumer, Home, Government and Enterprise) services. Focusing on the upgrade of automation and intelligence capabilities, the solution gradually realized the leap from partial automation to process automated closed-loop, and synchronously promoted the evolution of the operation and maintenance model from human-led to system autonomous closed-loop, fully supporting the transformation of network operation and maintenance towards unattended and active autonomy.

L4 high-value scenarios (2025~): This stage involves comprehensively laying out high-value scenarios, moving from capability maturity to the harvest period of commercial success. The core objective is to rely on mature end-to-end automated and intelligent closed-loop scenario construction solutions, driving large-scale implementation with quantifiable value effectiveness. By employing the standardized methodology of current status goal setting > end-to-end process sorting > three-layer AI full-stack evolution (including security capability embedding) > deployment verification, ZTE deeply integrates AI security capabilities into the full-stack evolution path, ensuring stable and reliable achievement of L4 autonomy in complex real environments. This integration path not only supports service value closed-loops, but also fully meets domestic regulatory requirements and international compliance standards such as the European Union AI Act, building a trustworthy, safe, and sustainable evolution system for large-scale AI implementation.

Based on operational value and technological maturity, and based on the 20 typical scenarios recommended by TM Forum, ZTE has expanded and selected the following 30+ high-value scenarios:

Service-oriented		Consumer Service			Home Service			Government and Enterprise Service						
		Voice	Internet access	...	Home broadband	FTTR	...	5G private network	IoT	Cloud access dedicated line	Cloud interconnect dedicated line	Internet dedicated line	Data dedicated line	...
Operation	Pre-sales support							★						
	Service provisioning				★	★		★		★	★	★	★	
	Network complaint handling	★			★			★	★	★				
	Service fault monitoring	★												
	Service quality optimization				★			★				★		

Network-oriented		Wireless Network (4G, 5G)			Core Network (EPC, 5GC, etc.)	Network Cloud (Server, virtual layer)	Transport Network (OTN, SPN, PON, etc.)	IP Network (CMNET, cloud private network, etc.)	Power&Environment (Power environment of the data center)	...
Planning	Network Resource Onboarding & Control	★								
Maintenance	Network fault monitoring	★			★	★	★	★		
	Network change monitoring				★	★	★	★		
Optimization	Network performance optimization	★	★	★			★	★		
	Network energy efficiency optimization	★	★			★	★		★	
	Resource utilization optimization						★			

Table 2-1 ZTE Autonomous Network L4 High-Value Scenarios

ZTE has defined the L4 roadmap by year: By the end of 2025, a total of 21 high-value scenario L4 pilot breakthroughs have been completed; by the end of 2026, a total of 30+high-value scenario L4 pilot breakthroughs will have been completed. At the same time, with service development and network evolution, more high-value scenarios of reducing costs, monetizing revenue, and improving perception will be further planned to essentially tackle L4 goals.

2.1.2 ZTE L4 High-Level Autonomous Network Solution Philosophy

Based on L4 pilot practices with global operators, ZTE deeply analyzes the core capability requirements and industrial evolution path of high-level autonomous networks, and systematically builds on ZTE's L4 high-level evolution solution system outlined by value-driven, layered intelligence injection, autonomous decision-making, and automated closed-loop. Through full-stack AI architecture reconstruction and value scenario expansion, ZTE promotes the paradigm leap of the network from improving operation and maintenance efficiency to increasing business revenue, realizes the upgrade of the network from passive operation and maintenance to active operation, and accurately implements the goals of improving quality and efficiency, reducing costs, and increasing revenue.

At the value-driven level, the KBI-KEI-KCI effectiveness system is used as the core traction to form a full-stage end-to-end value closed loop. Through quantitative indicators, optimization of all stages of network operation is forced to achieve endogenous cost reduction and efficiency improvement. At the same time, ZTE deeply expands high-value scenarios of market-network-service collaboration, transforms network capabilities into market revenue growth engines, breaks barriers between service and network, and makes the network a core asset empowering business growth.

At the technical capability level, ZTE builds a solid foundation of endogenous intelligence, builds an automation closed-loop base with network element real-time perception and single-domain/cross-domain collaborative remote control capabilities, and achieves reduced-manpower and unmanned operation in all scenarios through process reshaping. Superimposed with the AI innovation technology matrix, Nebula Telco Large Model, knowledge graph, Co-Sight agent factory, and digital twin technologies work together to create autonomous decision-making and automated closed-loop capabilities, accelerating end-to-end autonomous implementation. ZTE will continue to promote architecture evolution, build an Agentic AI architecture system with autonomous planning and collaborative capabilities, relying on an open, decoupled, separable, and combinable digital intelligence nebula orchestration architecture to adapt to diverse service needs and achieve autonomous service goals.

ZTE L4 high-level autonomous network evolution solution features value effectiveness driving. With the advantages of end-to-end automation and Agentic AI intelligent operation and maintenance, it not only ensures the efficient operation of the current network but also reserves flexible space for future technology iteration, achieving the continuous symbiotic evolution of business value and technical capabilities.

2.2 ZTE L4 High-Level Autonomous Network Evolution AIR Net Solution

ZTE's L4 high-level autonomous network evolution solution adopts a composable digital nebula architecture, with automation as the foundation and full-stack AI as the core. Anchoring the industry goal of Three Zeros and Three Selfs (Zero Wait, Zero Touch, Zero Trouble; Self-configuring, Self-healing, Self-optimizing), it accelerates the evolution of autonomous networks (AN) to the L4 high-level stage through the autonomous decision-making and automated closed-loop capabilities of agents/Intelligent-Digital Employee.

The solution adopts a layered intelligence injection architecture: The bottom layer consolidates the intelligent hardware foundation, the middle layer builds network-oriented agents, and the upper layer deploys service-oriented agents, supporting the progressive evolution of the Agentic AI architecture system. It allows the co-construction of an eco-system through open interfaces and has elastic scalability for future applications. At the same time, relying on the development platform and digital twin technology, it promotes product agent transformation with AI agents as the core grasp, building a Intelligent-Digital Employee system.

The solution not only realizes the commercial closed loop of saving people, time, money, and generating revenue, but also forms a standardized Solution Package, which is open and shared with the whole industry, providing a unified paradigm for the large-scale implementation of L4 autonomous networks.

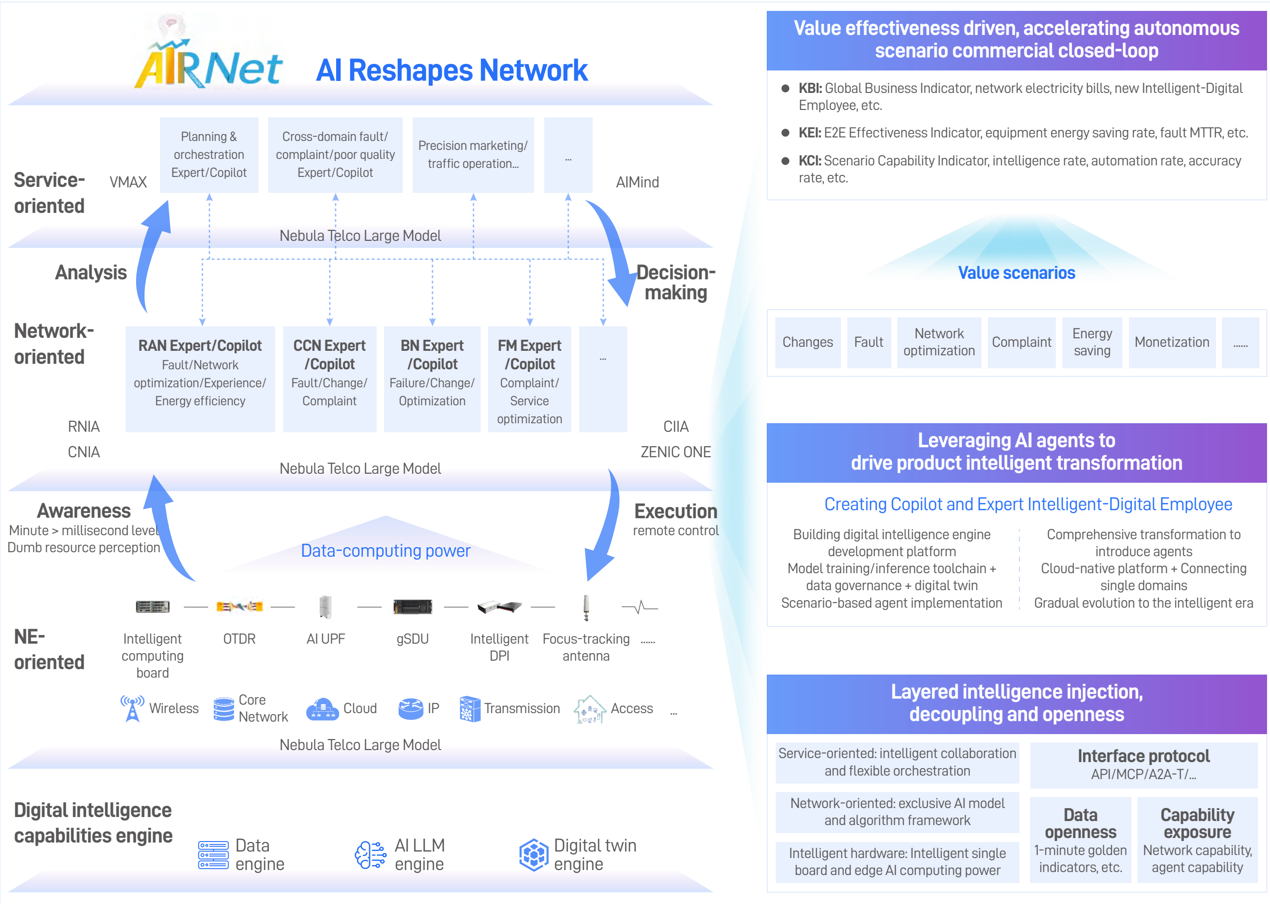

Figure 2-2 ZTE AIR Net Solution

The solution contributes practical experience to the industry from three dimensions:

I. Layered intelligence injection, decoupling and openness

Network element-layer: At the hardware atomic capability layer of network elements, ZTE embeds intelligent chips and deploys edge AI computing power modules, ensuring hardware is equipped with intelligent perception capabilities and remote execution capabilities.

Network-layer: For single-domain networks such as wireless, core, and transmission, ZTE builds exclusive AI models and algorithm frameworks, forms security verification mechanisms, and gradually realizes single-domain autonomous operation and maintenance.

Service-layer: ZTE creates cross-domain service intelligent collaboration, breaks barriers to inter-domain data and capabilities, orchestrates and flexibly schedules multi-domain network capabilities on demand, and supports service intelligence connectivity. It also ensures the credibility of the entire process of data transmission, capability invocation, and collaborative decision-making through the built-in security authentication mechanism of the A2A-T protocol.

At the same time, ZTE collaborates with industrial partners, opens data and capability invocation interfaces, implies joint building of a compliance protection system with ecological partners, and co-builds a digital intelligence ecosystem.

II. Leveraging AI agents to create two types of Intelligent-Digital Employee: Copilot and Expert, driving product intelligent transformation

ZTE focuses on the in-depth development of agent technology. On the one hand, it builds an intelligent engine development platform, integrates tool chains such as model training and model tuning, accelerates AI agent production iteration, combines data governance to provide a high-quality corpus for AI training, and combines digital twins to enable rapid implementation of agents in operation and maintenance scenarios. On the other hand, ZTE promotes the agent transformation of products in the whole domain, connects the agent capabilities of single-domain products based on the cloud-native platform, realizes functions such as cross-domain fault scoping and positioning, and gradually evolves the entire network operation and maintenance into the intelligent era.

Through this solution, ZTE creates two types of industrial-grade communication Intelligent-Digital Employee: Copilot and Expert. This transforms L4 autonomy from a technical concept into mass-producible, manageable, and value-added network productivity, giving silicon-based life responsibilities, goals, and collaboration.

1. Evolution from Intelligent Tools to Intelligent-Digital Employee

When operators currently evolve towards L4 high-level autonomous networks, the usual agent development methods fall into a triple dilemma: First, large-scale deployment is difficult, the development cycle of single-scenario agents is long, and they cannot adapt to the ever-changing needs of the network; second, security control is weak, lacking clear boundaries of rights and responsibilities, and operational risks are uncontrollable; third, adaptability to complex scenarios is poor, decision-making is fragmented when facing cross-domain faults, and accuracy is low. These pain points have become the core bottlenecks restricting the release of autonomous network value. Relying on AI LLM engine capabilities such as Co-Sight agent factory and Nebula Telco Large Model, ZTE quickly produces industrial-grade communication Intelligent-Digital Employee online. ZTE Nebula Communication Intelligent-Digital Employee (hereinafter referred to as Intelligent-Digital Employee) endows agents with clear job responsibilities and SLA requirements, fully utilizing technical capabilities such as agent and Traditional-Generative AI model cooperation, giving silicon-based life a sense of purpose and mission. The Intelligent-Digital Employee system effectively solves key challenges faced by traditional agents in communication network applications, such as low large-scale deployment efficiency, insufficient security, and limited adaptability to complex scenarios, providing reliable technical support for operators' networks to evolve towards L4 high-level autonomous networks.

ZTE focuses on high-value scenarios of high-level autonomous networks. Through Intelligent-Digital Employee upgrades for all products in the domain, it builds a complete Intelligent-Digital Employee product system, including two major categories: Expert Intelligent-Digital Employee and Copilot Intelligent-Digital Employee. Among them, Expert Intelligent-Digital Employee feature a high degree of autonomy, enabling them to independently execute complex tasks and form closed-loop automation for backend processes; Copilot Intelligent-Digital Employee focus on assisting operation and maintenance engineers to improve the work efficiency of frontline personnel in executing tasks.

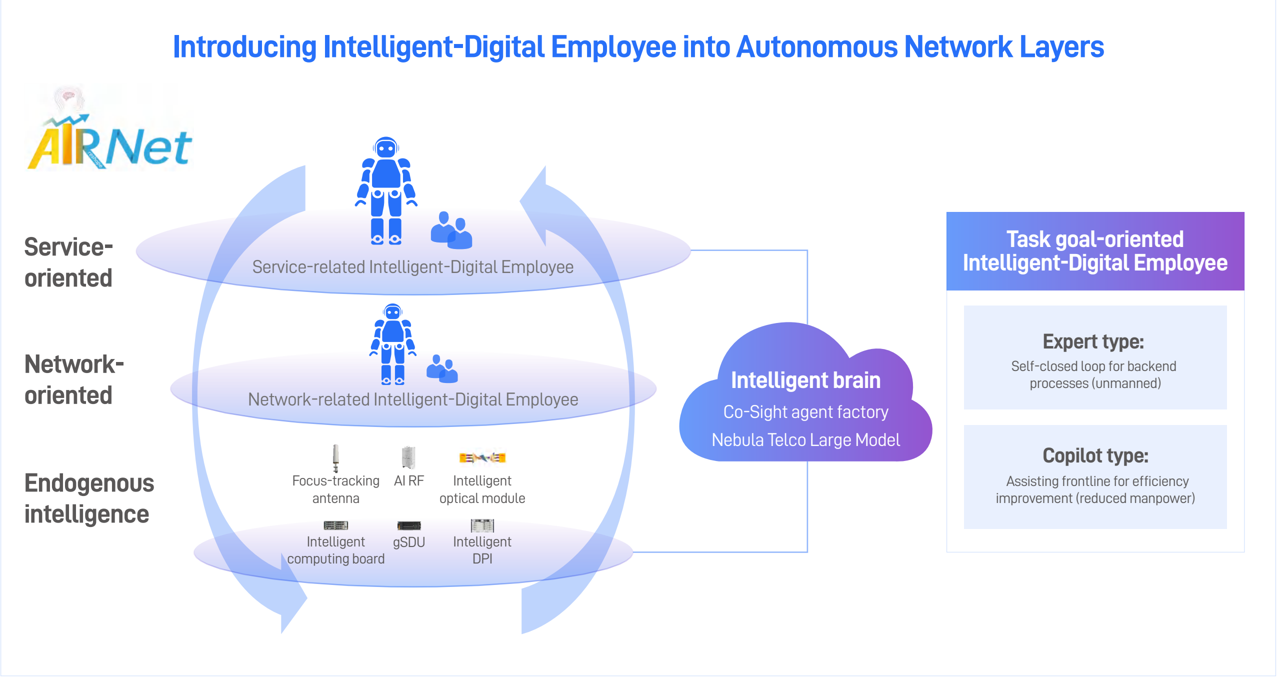

Figure 2-3 L4 High-Level Autonomous Network Intelligent-Digital Employee Solution

2. Defining four standards for industrial-grade Intelligent-Digital Employee

ZTE Intelligent-Digital Employee build four differentiated capabilities, setting standards for Intelligent-Digital Employee applications in the communication industry:

Goal-oriented and SLA-driven: Intelligent-Digital Employee adopt a job responsibility matrix and SLA) quantitative indicator system for standardized management, supporting task decomposition, collaboration, and iteration based on SLA requirements, and dynamically adjusting task scheduling policies based on real-time monitoring data to ensure stable achievement of end-to-end goals.

Multi-agent dynamic collaboration: Relying on the industrial-grade communication Intelligent-Digital Employee production platform of Co-Sight agent factory, it provides full life cycle management from development, deployment, to operation and maintenance. Through long-term task autonomous planning capabilities, precise matching between Intelligent-Digital Employee capabilities and scenario tasks is achieved; relying on multi-agent collaboration technology, it supports Intelligent-Digital Employee to organize tasks around goals, building a flexible command intelligent operation to ensure efficient collaborative execution of cross-domain multi-tasks.

Clarify security rights and responsibilities boundaries: Intelligent-Digital Employee build a role-based permission control system, dividing the operation boundaries and execution ranges of different types of Intelligent-Digital Employee. It has built-in full-process operation audit logs and intelligent security rollback mechanisms to achieve real-time monitoring of abnormal behaviors and trigger human intervention. Through a traceable rights and responsibilities control model, it solves security risks in the agent's autonomous decision-making process and ensures the compliant operation of Intelligent-Digital Employee in industrial-grade scenarios.

Continuous evolution and data closed-loop: Intelligent-Digital Employee build a learning closed-loop mechanism driven by data, such as online iteration and scenario generalization. At the same time, through automated collection and iterative analysis of evaluation cases, combined with strong reasoning model reinforcement learning frameworks for targeted optimization, Intelligent-Digital Employee can effectively alleviate the performance degradation problem caused by service model changes. Verification shows that compared with traditional agent solutions, task accuracy has significantly improved.

3. Practice philosophy: Anchoring the development path of new quality productivity

ZTE takes Three Self Capabilities as the implementation path and Three Zero Experiences as the development goal. Through full-stack AI technology, ZTE connects breakpoints in the operation and maintenance process, promoting a comprehensive transformation of network operation and maintenance from passive response to active prevention and from human-led to AI-led.

This practice not only builds a green, efficient, and sustainable new operation and maintenance system, but also drives network productivity upgrades through technological innovation, injecting strong momentum into the development of the digital economy and setting a replicable and promotable new paradigm for global high-level autonomous network construction.

III. Driven by high-value scenarios: Building a KBI–KEI–KCI value effectiveness system to create a commercial closed loop of efficiency, cost savings, and revenue growth, accelerating autonomous scenario commercial closed-loop

KBI: Global Business Indicator, network electricity bills, new Intelligent-Digital Employee, etc.
KEI: E2E Effectiveness Indicator, energy saving consumption, fault MTTR, etc.
KCI: Scenario Capability Indicator, intelligence rate, automation rate, accuracy rate, etc.

Below, we elaborate on the AIR net high-level autonomous network evolution solution from six core aspects: Value Orientation, Architecture Evolution, Endogenous Intelligence, AI Innovation, Process Reshaping, and Business Expansion.

2.3 Core Point 1: Value Orientation, Deepening KBI-KEI-KCI Effectiveness System

ZTE believes that from a value-driven perspective, the Autonomous Networks (AN) KBI-KEI-KCI effectiveness system is a value quantification framework featuring layered support and closed-loop linkage, which aligns closely with TM Forum's AN L4 level standards and the deployment needs of high-value scenarios. Its core logic follows the progressive relationship that capabilities underpin effectiveness, and effectiveness translates into value. The three indicators (KBI, KEI, KCI) aim to focus on value quantification, operational support, and capability enablement for AN L4 high-value scenarios.

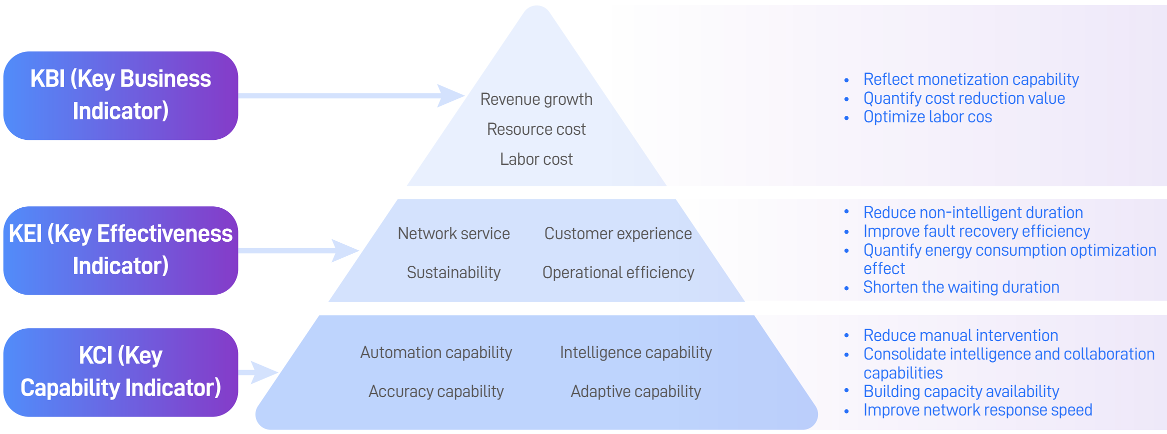

Figure 2-4 KBI-KEI-KCI Indicator System

KBI (Key Business Indicator): As the top-level value core of the system, it directly quantifies the actual contribution of Autonomous Networks (AN) to the business, focusing on four core areas: OPEX optimization, workforce efficiency improvement, business revenue growth, and risk control. Leveraging metrics such as OPEX reduction, time saved in network operations, business revenue growth, and reduction in major outages, it transforms the operational effectiveness optimized by KEI into direct commercial value and business benefits.

KEI (Key Effectiveness Indicator): It encompasses four core dimensions: network service, customer experience, sustainability, and operational efficiency. Leveraging metrics such as Mean Time To Recovery (MTTR), SLA compliance rate, equipment energy saving ratio, and service delivery time, it transforms the technical capabilities of KCI into tangible operational effectiveness, reflecting the precision and efficiency of network O&M, while ensuring the stability of customer experience and the green performance of sustainable operations.

KCI (Key Capability Indicator): Focusing on the core technical support capability of Autonomous Networks (AN), and centered on capabilities such as automation, intelligence, and accuracy, it builds the technical underpinning for AN operation by leveraging measurable metrics including task operation automation rate, end-to-end automation rate, and proportion of intelligent infrastructure, providing solid capability support for upper-layer effectiveness metrics.

Each value scenario combines different value directions, selects the corresponding calculation formula to complete value effectiveness calculation and evaluation, and takes quantitative results and processes as the basis to promote the practical advancement of effectiveness, providing theoretical basis and guidance.

Indicator Level	Scenarios	Metric Name	Core Formula	Value Description
KBI	Revenue growth	Service revenue increase amount	Σ (Labor cost) + Σ (Resource cost) + Σ (Network monetization cost)	Reflect the monetization capability brought by AN
	Resource cost	Operation and maintenance cost reduction amount	Σ (Labor cost) + Σ (Resource cost) + Σ (Network monetization cost)	Quantify AN direct cost reduction value
	Labor cost	Network operation personnel savings	Σ (Internal personnel savings + Optimization service savings + Outsourced personnel savings)	Optimize labor cost
KEI	Network operation and maintenance	Time savings	Σ Non-intelligent duration of each stage - Σ Intelligent duration of each stage	Reduce non-intelligent duration of single scenarios
	Customer experience	Fault repair rate	(1 - Intelligent fault recovery time/Non-intelligent fault recovery time) × 100%	Improve fault recovery efficiency
	Sustainability	Equipment energy saving rate	(1 - Energy saving mode consumption/Non-energy saving mode consumption) × 100%	Quantify equipment energy consumption optimization effect
	Operational efficiency	Resource deployment time savings	Non-intelligent resource allocation time - Intelligent resource allocation time	Shorten user waiting duration

KCI	Automation capability	Task operation automation rate	(Automated sub-tasks/Total sub-tasks) × 100%	Reduce manual intervention
	Intelligence capability	Intelligence rate	(Intelligent infrastructure count ÷ Total infrastructure count) × 100%	Consolidate intelligent capability foundation
	Accuracy capability	Single stage accuracy rate	(Capability determination accurate counts/Total determination counts) × 100%	Determine key capability availability
	Adaptive capability	Adaptation duration	Response start time - Environment change detection time	Improve network dynamic response speed

Table 2-2 Examples of Value Effectiveness Indicators

2.4 Core Point 2: Architecture Evolution, Building an Agentic AI Architecture System

Against the background of the deep evolution of AI technology towards intelligence and autonomy, Agentic AI has become a core development direction. Its core value is reflected in realizing the deep internalization of human experience and knowledge system, promoting AI autonomous takeover of service processes, and completing the advanced evolution of AI to the human-like thinking mode. ZTE's autonomous network takes this as its technical foundation and relies on key technologies such as Agentic AI, model self-evolution, multi-agent collaboration, and twin simulation to build the next generation of full-stack Agentic AN technology architecture.

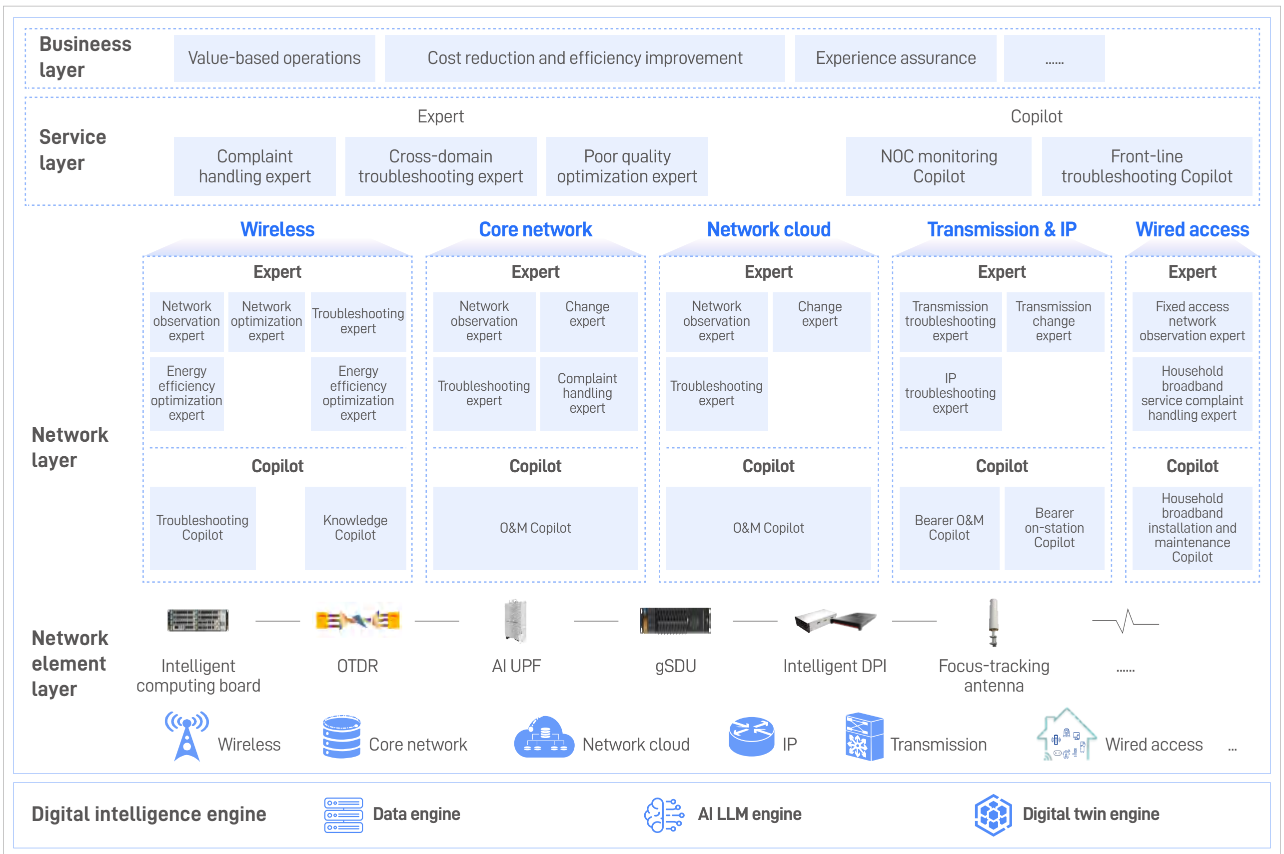

Figure 2–5 Agentic AI Architecture

As illustrated in the figure, we will briefly explain the Agentic AI architecture from bottom to top:

Digital intelligence engine: Through data engines, AI LLM engines, and digital twin engines, an intelligent network operation and maintenance system is built to help autonomous networks move towards a higher level of intelligence. Among them, the data engine enhances capabilities such as computing acceleration, multi-modal data lake, AI Ready data governance, and intelligent data analysis; the AI LLM engine enhances capabilities such as super agent framework, graph model platform, corpus synthesis tools, one-click fine tuning, and algorithm evolver; the digital twin engine enhances capabilities such as simulation, decision-making, and 3D modeling visualization.

NE-oriented: By adding smart network element hardware, it realizes local real-time decision-making and execution through native agents, and adapts to the cross-layer collaboration requirements of Agentic AI architecture, providing low latency and highly reliable underlying support for L4 autonomous networks and realizing "edge autonomy + global collaboration". In the future, we will gradually promote the standardization, lightweight, and security reinforcement of network element agents to achieve the integration and coordination of network elements from different equipment manufacturers.

Network-oriented: The wireless network provides network observation experts, network optimization experts, troubleshooting experts, energy efficiency optimization experts, experience assurance experts, as well as troubleshooting Copilot and knowledge Copilot. The core network provides network observation experts, change experts, troubleshooting experts, complaint handling experts, and O&M Copilot . The network cloud provides network observation experts, change experts, troubleshooting experts, and O&M Copilot . The transmission & IP network provides transmission troubleshooting experts, transmission change experts, IP troubleshooting experts, as well as bearer O&M Copilot and bearer station visit Copilot. The wired access network provides agents such as network observation experts, household broadband service complaint handling experts, and household broadband installation and maintenance Copilots.

Service-oriented: It provides such agents as complaint handling experts, cross-domain troubleshooting experts, poor quality optimization experts, as well as NOC monitoring Copilot and front-line troubleshooting Copilot. Relying on the multi-agent collaboration capability, it builds an intelligent service system, creates end-to-end full-chain service support capabilities, achieves full coverage of customer needs, and forms a service closed-loop of "where the demand is, the service is", laying a solid intelligent foundation for high-quality service operations. Looking to the future, we will further integrate network management and service operations to gradually form an integrated intelligent hub for network operations and maintenance.

2.5 Core Point 3: Endogenous Intelligence, Consolidating the Foundation of Closed-Loop Automation

2.5.1 Real-Time Perception Capability of Network Elements

The network element perception capability is a key link in the intelligent closed loop of network operation and maintenance, which directly affects the effectiveness of the decision-making of the upper-level network management system. The comprehensiveness of problem and anomaly identification, the real-time nature of data reporting, and the passive device status perception capacity are key issues that need to be solved at the network element level. To adapt to the development of autonomous networks, the NE-layer shall evolve its capability of perception in the following aspects:

Perception dimension: From network performance-oriented to application experience-oriented. Through the base station intelligent computing integration version and the intelligent UPF of the core network, the intelligent perception of wireless network services is realized. Through the built-in intelligent board of OLT/Bras equipment, the intelligent perception of fixed broadband network service data is realized. Through automatic sampling of service data and network element AI model reasoning, the relevant intelligent board realizes feature extraction of collected data and accurately identifies and distinguishes different types of application traffic, such as video streams, voice streams, and game streams in the network, to further realize user experience portraits based on KQI indicators and establish a comprehensive perception system based on multidimensional data.

Perception accuracy: From periodic low frequency to adaptive variable frequency acquisition. In the conventional 15-minute cycle performance collection mode, data lag often makes it impossible to capture network fluctuations and service degradation in time, making it difficult to meet the real-time requirements of automatic delimitation and localization of network faults and automatic optimization of service quality. The practice of simply shortening the data collection cycle will lead to an increase in transmission bandwidth and storage costs. Therefore, building adaptive variable frequency data collection capabilities has become an inevitable evolutionary direction. When the intelligent single board senses the sudden deterioration of user experience, it triggers the second-level/-millisecond-level high-frequency collection of network element data in the form of events, and when the abnormality is restored, it returns to the conventional data collection frequency, which can well resolve the contradiction between perception accuracy and operational economy requirements.

Perception level: From active to passive. Optical fiber dumb resources are invisible and unmanageable, which is a traditional problem of network operation and maintenance. With the help of technologies such as intelligent optical modules in optical transmission equipment and Light ODN optical fingerprint splitters, it can be effectively improved and solved. Through the feature identification, data collection and analysis of optical signals, key capabilities such as fiber topology restoration, abnormal point localization, and same trench and cable identification can be achieved, providing complete basic data support for closed-loop processing of network problems.

2.5.2 Remote Execution Control Capability of Network Elements

In the wireless network optimization scenario, to optimize the antenna coverage direction and coverage area, it is necessary to adjust such parameters as antenna position, azimuth, and downtilt angle; however, the outfield mechanical antenna cannot support remote adjustment capabilities, and only on-site operation and maintenance engineers can be arranged to go to the station for adjustment, but the entire process of station visit is time-consuming and of high cost in operation and maintenance.

To improve the efficiency of network operation and maintenance and optimization, reduce the station-visit number of FME engineers, increase the automation rate of service processes, and reduce workflow processing delays, it is necessary to introduce intelligent hardware such as focus-tracking antennas, and enable their control via remote commands.

ZTE's intelligent hardware, such as focus-tracking antenna and gSDU (Green Site Digitalized Unit), has remote control capabilities, which can effectively improve the automation rate of network optimization and energy saving, improve customer experience and energy efficiency indicators, and further reduce operation and maintenance costs.

Scenario 1 Long-range performance optimization of focus-tracking antenna

Process: Based on the remote inclination/azimuth/beam width tuning process of user MR and based on the user's mobility characteristics, realize the precise delivery of network resources with "waves following people".

Diagnosis: The optimization agent analyzes the MR and cell wireless performance data reported by users and finds that the distribution of users in a specific geographical area shows the characteristics of tides. In some periods, there are problems with coverage and throughput caused by the concentration of users on edge coverage.

Policy generation: The optimization agent determines the current performance problem based on the AI algorithm and can adjust different antenna azimuths, downtilt angles, and wave widths in multiple time periods to achieve automatic adaptation to changes in the tidal distribution of users.

Remote instructions: The optimization agent remotely sends adjustment instructions for antenna RET electrical downtilt, RAS azimuth, and beam width at different time periods to the control unit of the focus-tracking antenna through the target network element RRU.

Real-time monitoring and adaptive rollback: After the focus-tracking antenna executes the instruction, the optimization agent monitors the KPI changes in the area in real time. If it is found that the adjusted KPIs (such as throughput or SINR) are not improved or deteriorated, the optimization agent will immediately trigger a remote recovery operation to roll back the configuration to the previous stable configuration point without manual intervention.

Application results: The application of focus-tracking antennas realizes the optimized "station-visit-free, full remote, and automatic" autonomous capabilities for coverage and performance optimization. The time spent on a single RF tuning is shortened from 2 hours in the traditional mode to less than 10 minutes, with the efficiency increased by more than 10 times. Measured data from multiple sites show that after the "focus tracking" tidal optimization policy is enabled, the network performance and user experience in the pilot area are significantly improved. The MR coverage is increased by 1.27%, and weak coverage areas are improved significantly. The average daily total 4/5G traffic in the area is increased by 2.35%, and the network experience lag rate is decreased by more than 10%.

Scenario 2 Automatic starting and stopping of load wake-up based on gSDU

Process: As shown in the figure, when the energy-saving time is up, evaluate the load and perception KPI. If the evaluation result is good, the reliability judgment is passed, the user is migrated, the corresponding port of gSDU is powered off, and AAU/RRU is powered off. During the energy-saving interval, continue to evaluate the load and perception KPI. If the load increases or the perception deteriorates, the corresponding port of gSDU is powered on, and AAU/RRU wakes up.

Application results: The automatic starting and stopping of AAU/RRU based on load wake-up is realized to achieve "0 load and 0 consumption" during off-peak hours and minute-level wake-up during peak hours to ensure experience and guarantee energy saving and perception.

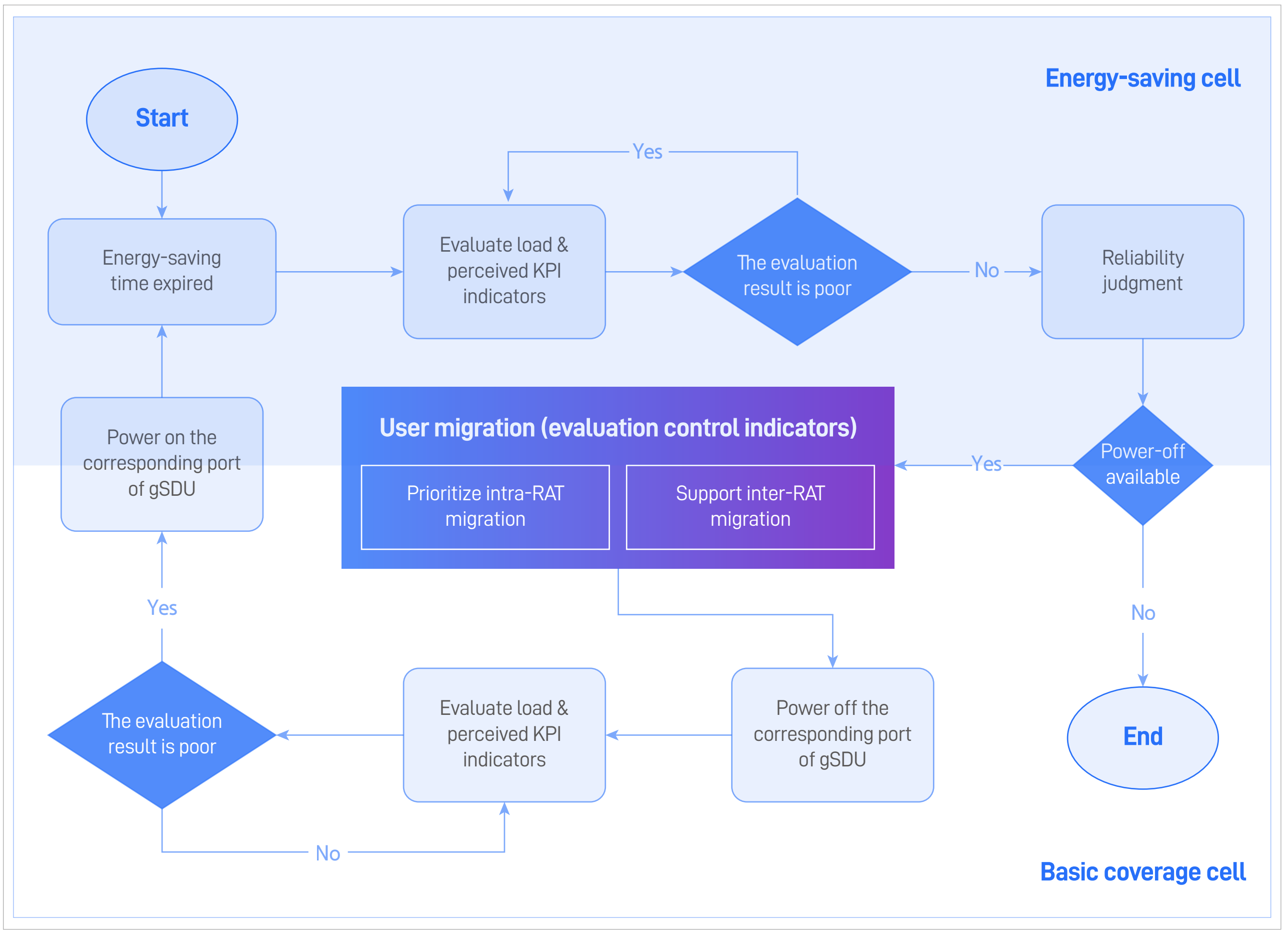

Figure 2-6 Principle of Extreme Sleep Based on Load and Perception

Scenario 3 Automatic diagnosis of fronthaul link faults

Process: Automatically perform fault diagnosis and root cause analysis based on alarm data.

Alarm: The intelligent optical module continuously detects the received optical power. If the current transmission link fails and causes the optical path to be interrupted, the system reports a disconnection alarm.

Remote diagnosis: The troubleshooting expert automatically starts the OTDR (optical time domain reflectometer) function integrated with the intelligent optical module remotely according to the alarm information, and realizes accurate geographical localization of the fault point by transmitting detection light signals, receiving and collecting reflected light information, and combining intelligent analysis algorithms.

Root cause analysis: The troubleshooting expert gives the root cause corresponding to the alarm based on the fronthaul link topology information and remote diagnosis information.

Application results: Through remote automatic diagnosis and root cause analysis of the fronthaul link, intelligent task routing and rapid resolution of fronthaul link faults are achieved to avoid repeated station visits and improve network service time.

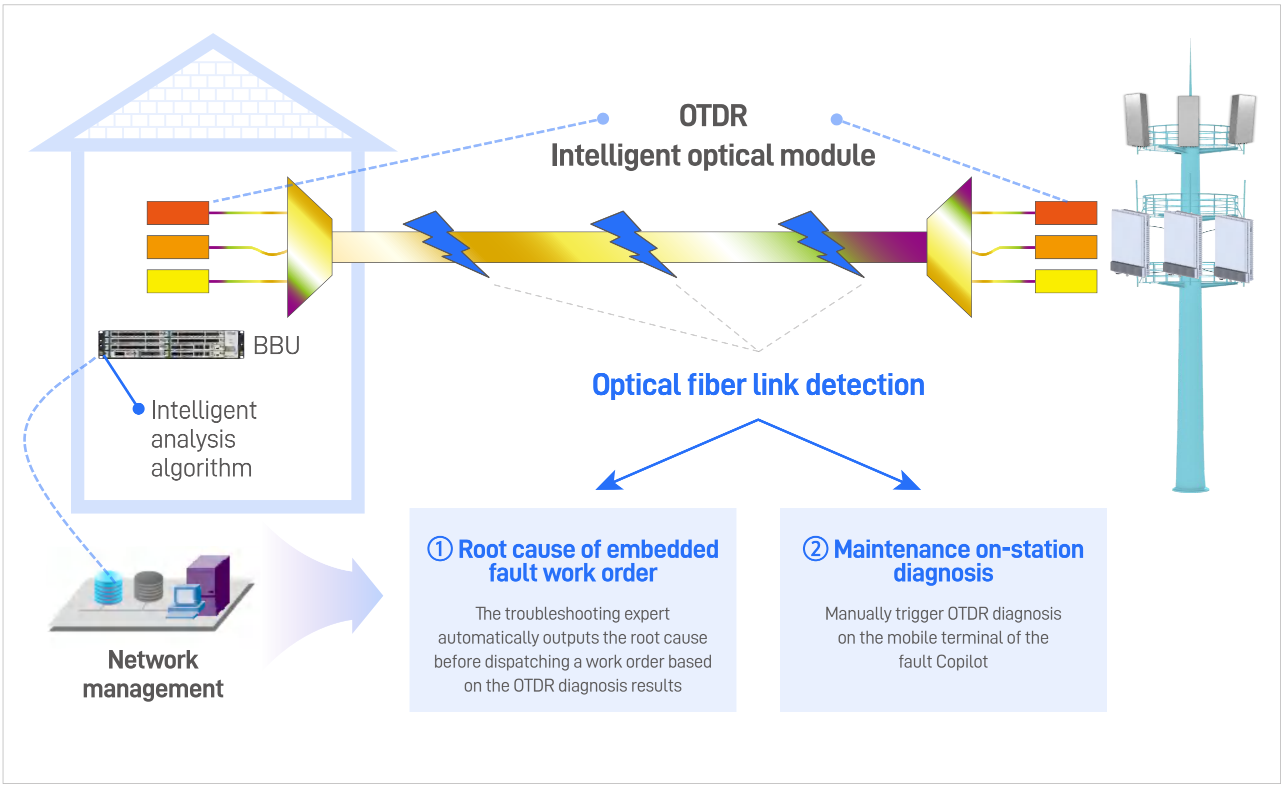

Figure 2-7 Precise Identification and Positioning Fronthaul Link Fault of OTDR Intelligent Optical Module

2.6 Core Point 4: AI Innovation, Building Independent Decision-Making Capabilities

ZTE's Digital Intelligence Engine is the cornerstone of autonomous networks. It is driven by three core engines as described below: the data engine builds a full-link high-quality data base, not only providing real-time data support, but also embedding graph model platform capabilities, and through Schema-oriented automated knowledge graph construction and GraphRAG graph hybrid retrieval reasoning technology, it realizes accurate extraction of entity relationships of multi-source heterogeneous data and logical connection of discrete knowledge points, providing deep knowledge support for decision-making; as an intelligent hub, the AI LLM engine realizes the capability transition from language understanding to autonomous task execution, and; the digital twin engine builds an accurate network mirror to provide simulation verification and trusted guarantee for decision-making.

Based on the digital intelligence engine, with the Nebula Telco Large Model as the reasoning and decision-making brain, together with the professional cognitive enhancement of the graph model platform and the simulation verification foundation of digital twin networks, through the Co-Sight agent framework and agent factory, various professional network-oriented agents (such as faults, network optimization, energy efficiency) and service-oriented agents (such as complaints, poor quality, service orchestration) are efficiently produced throughout the "perception-analysis-decision-execution" full process closed loop. Based on the collaboration between agents through A2A-T and seamless interaction with the tool system through MCP, a flexible and scalable multi-agent collaboration network is developed. This architecture breaks the empirical shackles of "manual dominance and single-domain fragmentation" in traditional operation and maintenance. Through real-time perception, data-driven, cross-domain collaboration, proactive prevention, and multi-target optimization, it achieves a paradigm upgrade from "merely reacting to problems" to "autonomously completing complex tasks".

We build an Agentic framework with agents as the core to break the limitations of fragmented AI through multi-agent collaboration, and achieve end-to-end collaboration through Nebula Telco Large Model + graph model (knowledge graph & graph retrieval) + digital twin + intelligent hardware, support autonomous decision-making by agents to support end-to-end automatic closed loop. In detail, strong reasoning and hybrid reasoning models are used to build Co-Sight agent factories and model evaluation sets to improve the accuracy of long-tail problem solving. Graph models (knowledge graphs & graph retrieval) are developed to improve result stability. Digital twins + policy pre-validation are applied to improve operational safety.

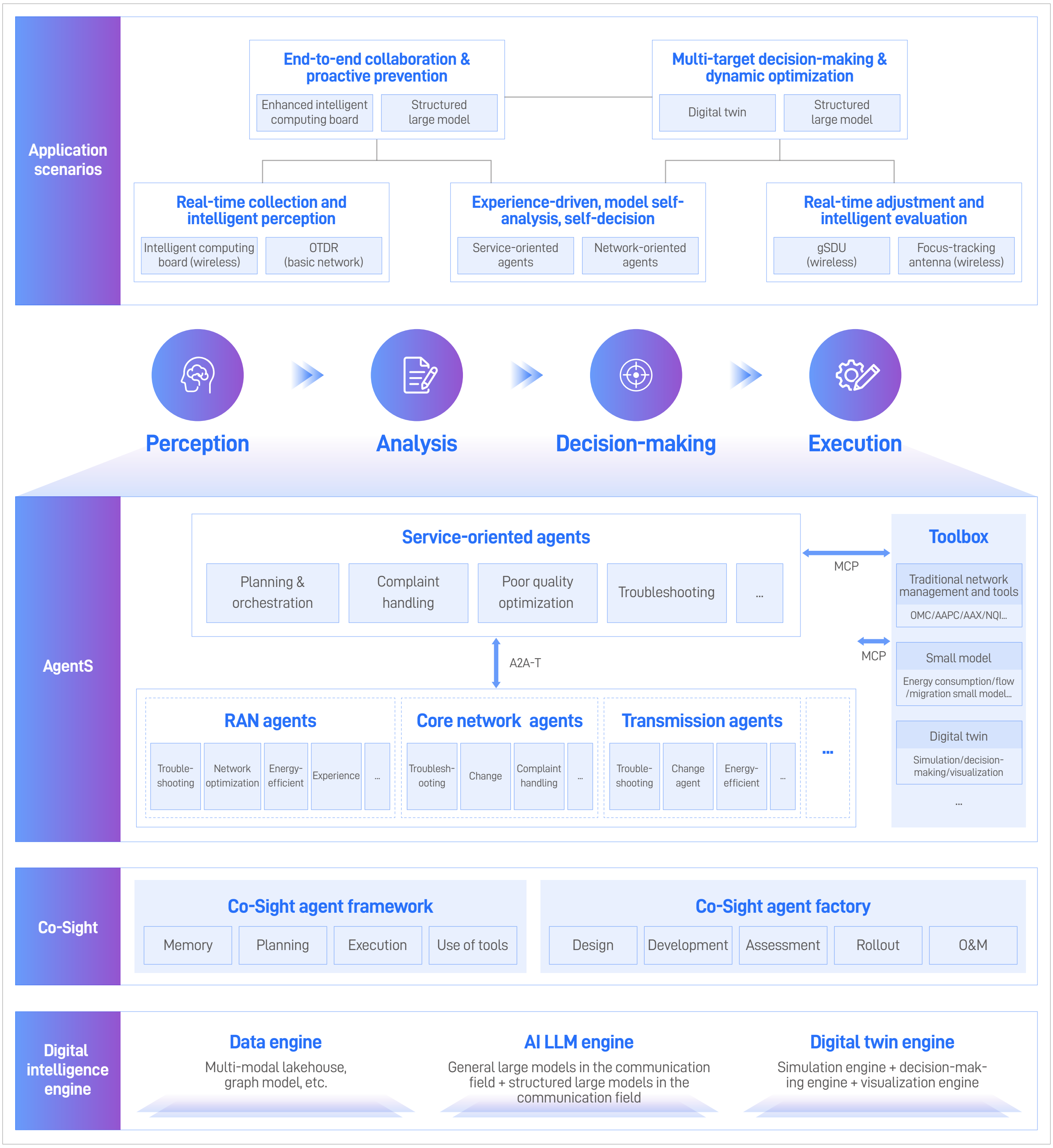

Figure 2–8 Agent Innovation Empowerment Framework

2.6.1 Nebula™ Telco Large Model

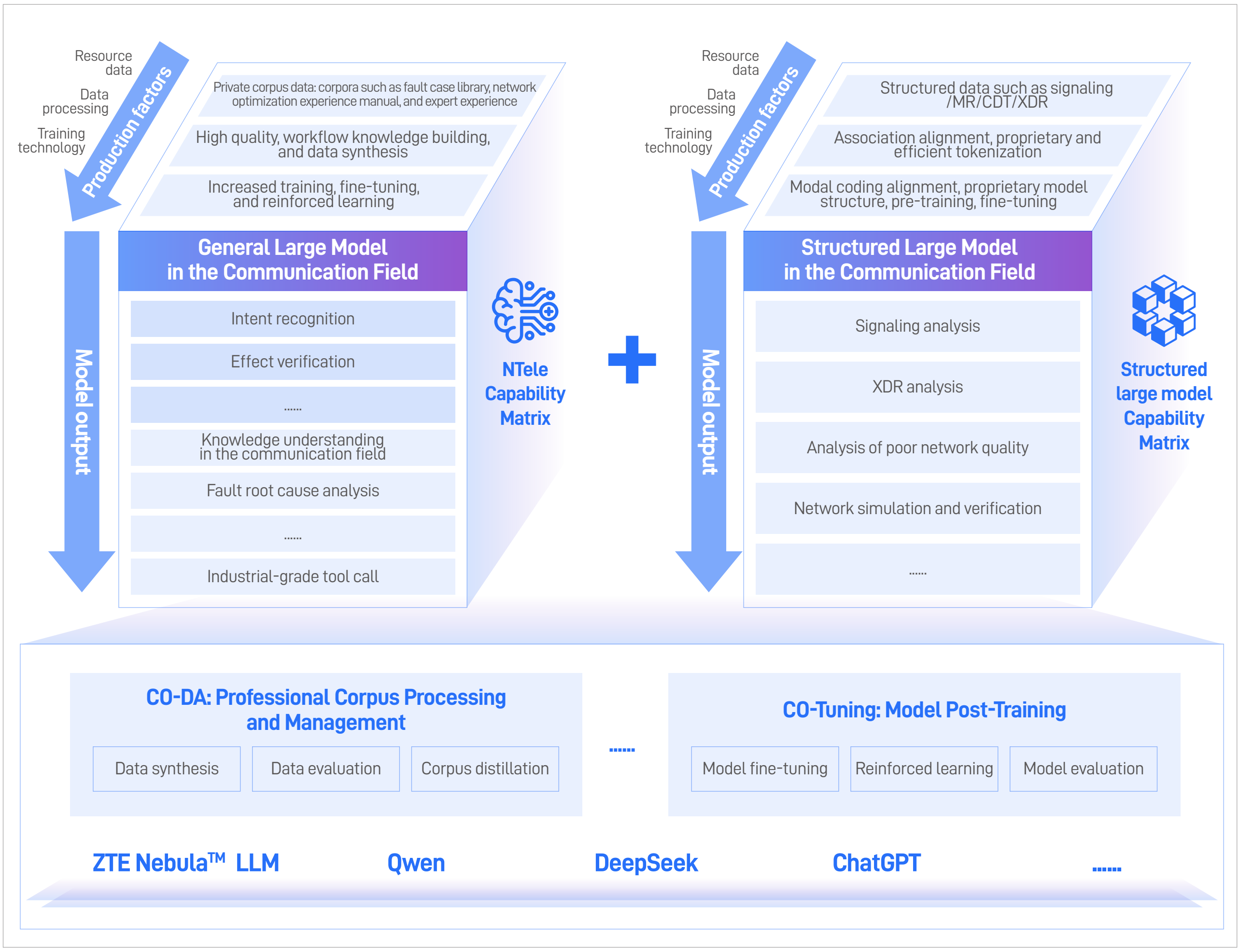

Figure 2-9 Architecture of ZTE Nebula™ Telco Large Model

In the process of moving towards L4 high-level autonomous networks, operators face complex operation and maintenance challenges intertwined with massive unstructured documents, user interaction intentions, and underlying signaling logs. General large models—despite their broad capabilities—are inherently limited in meeting the communications industry's stringent requirements for professional depth, real-time performance, security, and multi-vendor interoperability.

To address these gaps, ZTE has developed Nebula™ Communication LLM: not as a standalone foundation model, but as a cohesive, operator-deployable system that integrates three tightly coupled components:

- a communication-domain generalist large model (NTele) for high-level policy reasoning, intent understanding, and knowledge synthesis;
- a lightweight structured large model specialized in real-time processing of network data (e.g., signaling, logs);
- and a secure, multi-vendor-capable AI engineering pipeline powered by the CO-DA and CO-Tuning platforms.

This "1+N" model matrix operates through a brain-cerebellum collaborative mechanism, enabling end-to-end intelligence that is deeply adapted to communication scenarios. As a result, Nebula™ Communication LLM has significantly improved the accuracy of solving long-tail problems in complex environments, while its integrated engineering system ensures robust data governance, efficient training, and secure deployment across heterogeneous operator networks.

I. 1+N Model Matrix

1. General large model in the communication field (NTele)

NTele serves as the communication-domain generalist engine of the ZTE Nebula™ Communication LLM.It innovatively adopts a hybrid reasoning architecture and focuses on reasoning and decision-making for complex service logic. Based on the infusion of massive operators' private high-value corpus (such as fault cases, operation and maintenance procedures) and digital twin synthetic data to internalize expert experience, this model deeply integrates the "fast thinking" capability of general generation and the "slow thinking" capability of logical strong reasoning. Compared with the general open source model, it not only has deep industry knowledge, but also can handle long-chain complex problems through a deep chain of thought (CoT). In key scenarios such as core networks and bearer networks, this hybrid capability enables it to demonstrate more accurate service intent analysis, complex tool chain calls, and complex fault root cause localization capabilities, which can effectively break through the dual limitations of general models in professional depth and logical breadth.

2. Structured large model in the communication field

The model is specially designed for processing structured network data, such as signaling and logs. It adopts a lightweight architecture to overcome the bottleneck of real-time processing of massive data and realize accurate modeling of the underlying network status. In high-frequency scenarios such as anomaly detection and user portrait construction, the model can support millisecond-level response on the edge side at a very low computing power cost, releasing the underlying data value.

The general large model is responsible for high-level policy generation, and the structured large model provides underlying status verification. They are deeply coupled through a unified mechanism to build a dynamic closed loop of "high-level decision-making > underlying verification > feedback optimization".

At the model level, we innovatively construct a "brain-cerebellum collaborative evolution" mechanism, where: ① Brain (cognition) is responsible for overall coordination and general analysis capabilities; based on the general large model (NTele) in the communication field, it focuses on offline, deep logic difficult sample tackling and knowledge production, providing power for continuous evolution of the cerebellum; ② Cerebellum (execution) focuses on accurate problem discovery, solution generation and twin simulation drills in specific fields, and, using lightweight structured large models, focuses on real-time, highly-concurrent anomaly detection and root cause classification to ensure more accurate discovery of root causes of poor quality, and realize applications in scenarios such as wireless network optimization and network planning. It is expected that the identification of poor quality root causes will be more accurate (↑ 30%), the model capability update cycle will be faster (↓ 50%), and the data labeling efficiency will be higher (↑ 50%).

Looking ahead, the evolution toward a communication-specific "world model" with capabilities such as counterfactual reasoning and implicit dynamics understanding will follow a progressive, software-driven roadmap—ensuring full backward compatibility and protecting operators' existing investments. In the near term (2026–2027), we will incrementally integrate digital twins with generative simulation technologies within the current Nebula™ LLM framework to enhance specific high-value scenarios, such as network planning, fault replay, and what-if analysis. These capabilities will be delivered as software upgrades on the existing AI engineering system—including the CO-DA and CO-Tuning platforms—requiring no forklift replacement of deployed infrastructure. Building on this foundation, in the mid to long term (2028 and beyond), as generative AI efficiency and edge computing mature, we aim to realize a full world model that not only internalizes structured knowledge like communication protocols, channel characteristics, and network topology, but also transcends pure data-driven paradigms to enable millisecond-level forward-looking spatial deduction and direct output of optimal control policies. As the key enabler of L4 autonomous networks, this evolution will retain the dual-loop architecture: in the non-real-time domain, large models and high-fidelity virtual sandboxes will perform long-horizon counterfactual reasoning and knowledge distillation; in the real-time domain, lightweight endogenous policies or distilled AI algorithms will execute millisecond-level decisions. This "cognitive-driven" paradigm will thus emerge incrementally, overcoming traditional bottlenecks of latency and energy consumption while delivering full-dimensional perception, self-configuration, self-repair, and self-optimization—laying a solid, future-proof intelligent foundation for 6G networks that seamlessly integrate sensing, computing, and intelligence.

II. CO Capability Platform

1. CO-DA: Copilot Data

CO-DA serves as the cornerstone for the ongoing advancement of model capabilities, dedicated to converting multi-source and heterogeneous raw industry data into high-quality knowledge assets. The platform integrates enhanced corpus synthesis tools and builds an end-to-end governance system from high-quality corpus generation and evaluation to corpus distillation. Through a multi-level verification mechanism and full-cycle management, it ensures the accuracy, compliance, and high value of the training corpus under the premise of strictly following data security specifications.

2. CO-Tuning: Large Model Post-Training Platform

The post-training platform CO-Tuning focuses on the continuous optimization and scenario adaptation of models. The platform covers three core capabilities for models: one-click fine-tuning, reinforcement learning, and evaluation and verification. It relies on the exclusive evaluation set built in the communication field to achieve model capability evaluation, while deeply injecting industry norms and security policies into the model. By establishing strict boundary constraints and real-time firewall mechanisms, the platform not only drives the evolution of model performance, but also ensures that it can achieve highly reliable and trustworthy operation in key telecommunications infrastructure with "no downtime in operations and no conflicting instructions".

For the Nebula Telco Large Model, ZTE adheres to the technical route of "integrated research and application" and has published nearly 20 core papers at top artificial intelligence academic conferences (such as ACL, AAAI, and ICLR), forming a full-stack self-developed innovation system. The "course reinforcement learning" algorithm accepted by EMNLP significantly improves the model's complex logical reasoning capability; the "multi-agent interaction protocol Co-TAP" recognized by IJCAI provides key support for the brain-cerebellum collaborative architecture, and the "TN-AutoRCA" communication root cause analysis framework included in AAAI effectively enables the precise localization of network faults. The above-mentioned cutting-edge academic achievements have formed ZTE's characteristic algorithms and are embedded in the core architecture of the Nebula Telco Large Model. They have all achieved engineering application transformation, established a technical benchmark for products in the industry, and won recognition from the industry.

2.6.2 Graph Model Platform

The graph model platform is responsible for the precipitation and dynamic update of structured, traceable, and verifiable knowledge. As a trusted external memory bank of large models, it provides an interpretable inference evidence chain to support compliance verification of high-risk operations. The graph model platform includes two core capabilities: knowledge graph building and updating, and hybrid retrieval reasoning, to support the construction of a full-link knowledge system from enterprise multi-source data to service intelligence applications.

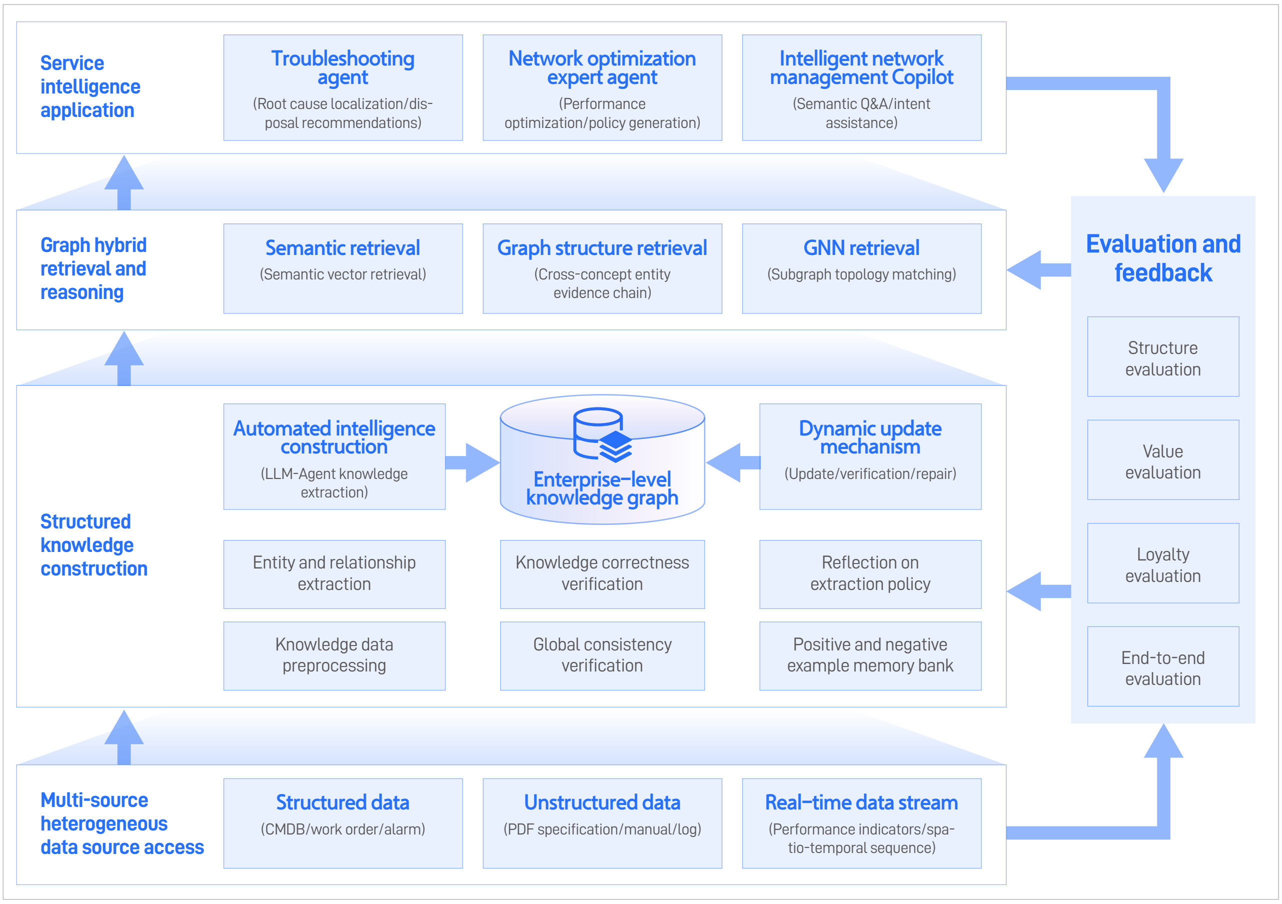

Figure 2-10 Graph Model Platform

Knowledge graph construction: ZTE's knowledge graph technology breaks through the traditional static construction mode and forms a dual-path construction of "static extraction + dynamic update". It also innovatively builds a self-checking and reflective evolution framework for knowledge extraction. Through automatic identification and logical backtracking of extracted Badcase, the correction strategy and new knowledge are synchronously deposited into the memory bank to achieve continuous self-optimization of the knowledge graph.

Graph hybrid retrieval: It is difficult for traditional retrieval to cope with complex multi-hop retrieval reasoning problems. Graph hybrid retrieval uses the semantic connectivity of graph structures to achieve logical connection of discrete knowledge points, effectively eliminating the reasoning faults caused by information fragmentation. Through the constraints of graph structures, the semantic relevance of LLM context is greatly enhanced, achieving a leap from "static matching" to "dynamic cognitive reasoning" and accurately capturing deep implicit evidence across documents.

Service practice and value embodiment: The above technologies have formed a complete application closed-loop in such scenarios as network optimization, intelligent operation and maintenance, and fault processing of China Mobile Innovation Laboratory. Taking China Mobile Innovation's fault handling as an example, the graph model platform automatically builds a fault propagation graph using massive historical work orders, fault manuals, and expert experience. When a new fault occurs, the system outputs a highly correlated fault context through graph hybrid retrieval technology, providing a factual basis for large model reasoning, effectively reducing illusions, and improving result stability. This solution has been applied to production in Changping and Chaoyang, Beijing, achieving an accuracy rate of more than 90% for wireless fault root cause localization and shortening the maintenance fault handling time by 10%, which strongly proves the advancement and commercial value of the technology.

2.6.3 Key Agent Technologies

I. Agent technology

Against the background of the deep integration of AI technology and communication networks, large models and agent technology are developing rapidly. As a core component of autonomous networks, agents are moving from theoretical concepts to large-scale applications, which promotes the transformation of network operation and maintenance from passive response to active prevention and autonomous decision-making. Agent technology has experienced a leap from simple conversation to autonomous execution. As AI technology enters the "landing period", autonomous network agents are also facing transformation. Copilot and Expert are reshaping the XOPS interaction model, and operation and maintenance have shifted from "people + platform" to "autonomous closed loop of agent service".

ZTE's independently developed Co-Sight agent framework is the first in the industry to propose full-link trusted computing. It has achieved a comprehensive upgrade of the traditional AI Agent framework by planning and executing a two-layer agent dynamic scheduling mechanism. It is an industrial-grade reliable agent development framework designed for high-complexity enterprise-level mission scenarios.

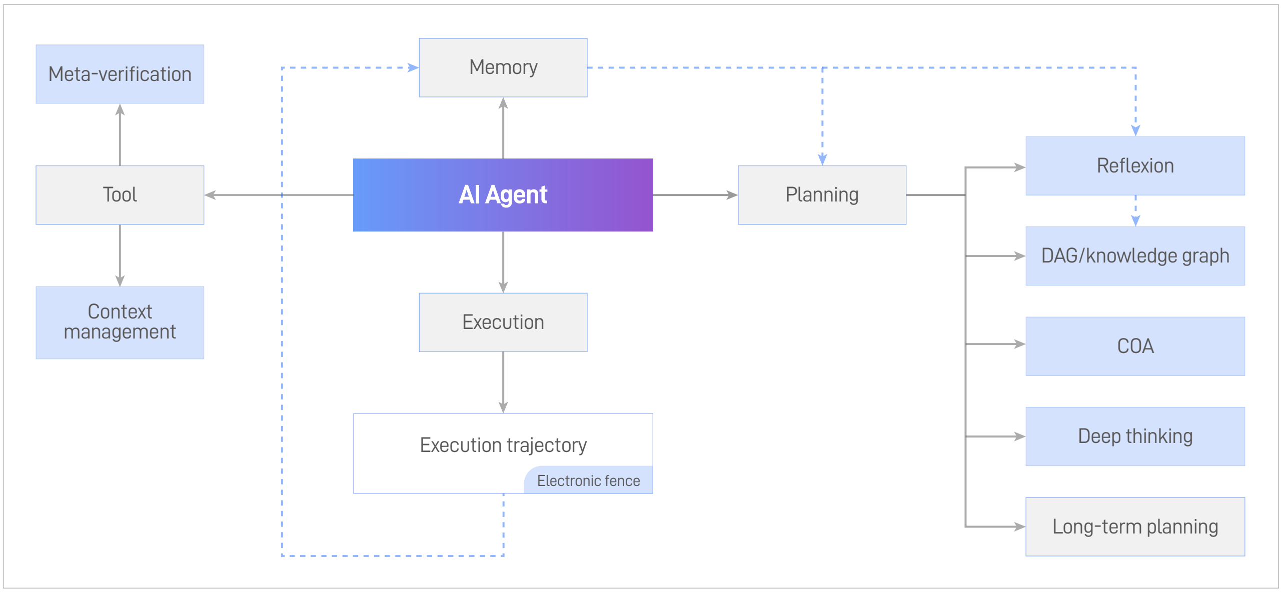

Figure 2-11 Co-Sight Agent Framework

The Co-Sight agent framework has the following technical features:

Cognitive enhancement: It achieves faster and more accurate intelligent decision-making. Through the separation of long-term planning and execution, deep reflection, concurrent thinking, and secondary planning mechanisms, it significantly improves the complex problem-solving capability and decision-making efficiency of agents.

Memory enhancement: It optimizes the action path to achieve precise execution. Agents can review and learn from past successes and failures like humans, and dynamically evaluate and optimize future action plans to continuously approach the best solution and avoid repeating mistakes.

Quality and safety: By setting up an electronic fence, it reviews unsafe processes and execution deviations.

Trusted computing: With meta-verification technology and structured contextual trusted reasoning technology, it verifies the reliability of information to avoid reasoning bias and insufficient exploration and improve the reliability of acquired information and derivation results, thus preventing error propagation from the source.

Based on the Co-Sight agent framework, the Co-Sight super agent has been realized and has been deeply commercialized in core scenarios such as operator autonomous network fault diagnosis and deep insight of intelligent computing integrated machines, significantly improving the level of operation and maintenance intelligence and decision-making efficiency. In China Mobile's troubleshooting agent scenario, the MTTR is reduced by 21.34%. The technical capabilities of this framework have been recognized by international authorities and continue to rank at the top of the world's top AI Agent lists, such as GAIA and HLE.

II. Co-Sight Agent factory

With the exploration of high-value scenarios in autonomous networks, operators have put forward higher requirements for quickly building autonomous agents and accelerating the launch of new services. However, the current construction of autonomous network agents faces the following challenges:

- Scenario fragmentation and difficulty in capability reuse: Network operation and maintenance scenarios are highly discrete (such as fault localization and capacity analysis), and each scenario requires independent development of intelligent algorithms and processes, resulting in duplicate construction and waste of resources.
- Long iteration cycle and slow value release: Agent development relies on manual design rules, and it usually takes several months from demand analysis to deployment.

ZTE's Co-Sight agent factory is a basic platform for quickly building agent applications, providing lean production intelligent factories, and helping to efficiently create industry-grade reliable agents. The agent factory provides a platform for building, evaluating, publishing, and running agents to help users quickly create, manage, and use agents. It provides different modes of agent production methods, which can be applied to the development of agent applications in various scenarios. Based on the concept of "industrial production", it transforms the development of agents from "manual customization" to "standardized assembly line", realizing the scale, configurability, and continuous evolution capabilities of agents.

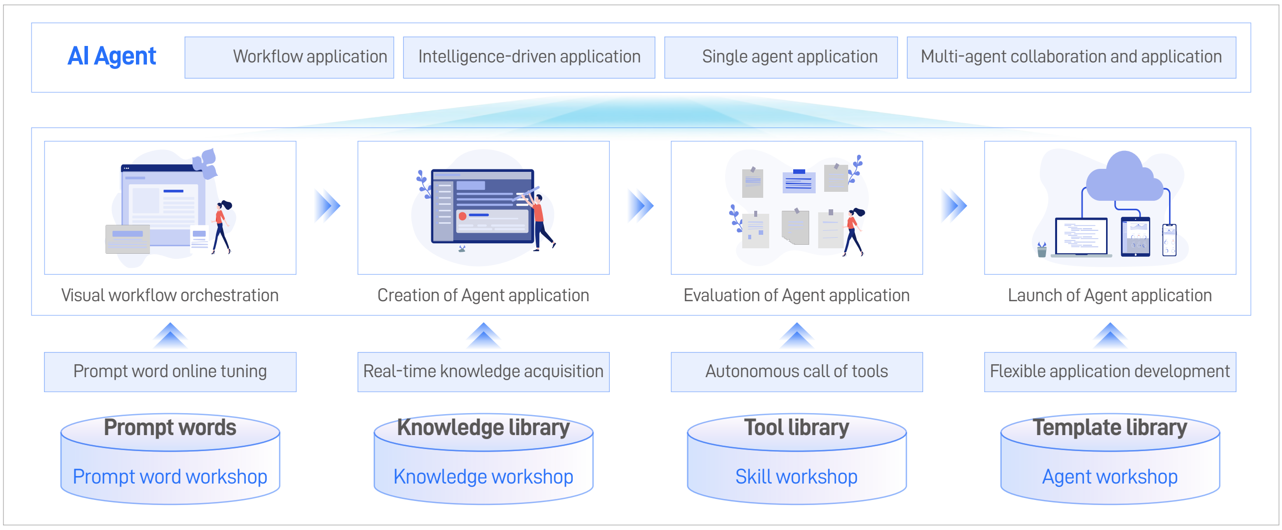

Figure 2-12 Co-Sight Agent Factory

III. Multi-agent Collaboration

The modern agent ecosystem is highly heterogeneous. Agents are developed using different frameworks (such as LangChain, AutoGen, and CrewAI), run on different platforms, and communicate using different protocols, such as the inter-agent protocol A2A, the message-oriented middleware communication protocol MCP, and the proprietary RESTful API. In the absence of a unified communication standard, developers have to create costly and fragile point-to-point integrations for each new agent or service. The resulting N-to-M adaptation problem not only hinders scalability, but also increases the vulnerability of the system. In other words, any change in an agent protocol may trigger a global cascade effect. The recent emergence of multiple protocols has further highlighted the urgent need for unified open standards.

The Unified Agent Protocol (UAP) launched by ZTE, with "modular architecture and ecological symbiosis" as its core concept, is committed to addressing the multiple challenges faced by multi-agent systems when deployed on a large scale, including lack of self-discovery capabilities between agents, difficulty in coordinating complex tasks, high adaptation burden caused by the coexistence of multiple protocols, and waste of resources caused by unnecessary contextual data transmission. The UAP decomposes multi-agent collaboration capabilities into independent functional modules such as AI gateway, registration and discovery, and authentication and communication protocols by establishing a unified multi-protocol service registry and gateway. Its main goals are to build a "proxy Internet", break the information silos between agents, provide a unified service plane and governance capabilities, support mainstream proxy protocols (such as A2A, ACP, and MCP), and lower the application integration threshold through protocol standardization and unified adaptation. It strives to strike an effective balance between system reliability, scalability, and security while accelerating application deployment. Its key goals are as follows:

Protocol integration: Develop a "universal language" for agent communication. The registration and discovery mechanism compatible with multiple mainstream protocols, such as A2A and MCP, is ZTE's feature, which has consolidated the interconnection foundation of agent networks.

Cross-domain penetration: Build a borderless collaborative working network, rely on high-performance gateway technology to achieve direct connection of agents across physical or logical domains, and provide efficient collaboration channels for wide-area distributed agent clusters.

Protocol adaptive conversion: Assist the smooth evolution of existing systems, with the help of multi-protocol intelligent conversion capabilities at the gateway layer, such as protocol conversion between A2A-T, A2A, ACP, and MCP, standardize system access methods, and support stock and incremental agents to participate in the new generation of collaborative systems without reconstruction.

Built-in zero trust mechanism: Build a solid cross-domain interaction security line, implement the zero trust security concept, carry out two-way identity authentication and session encryption, eliminate unauthorized access and man-in-the-middle attacks, and ensure that every cross-domain collaboration is traceable and auditable.

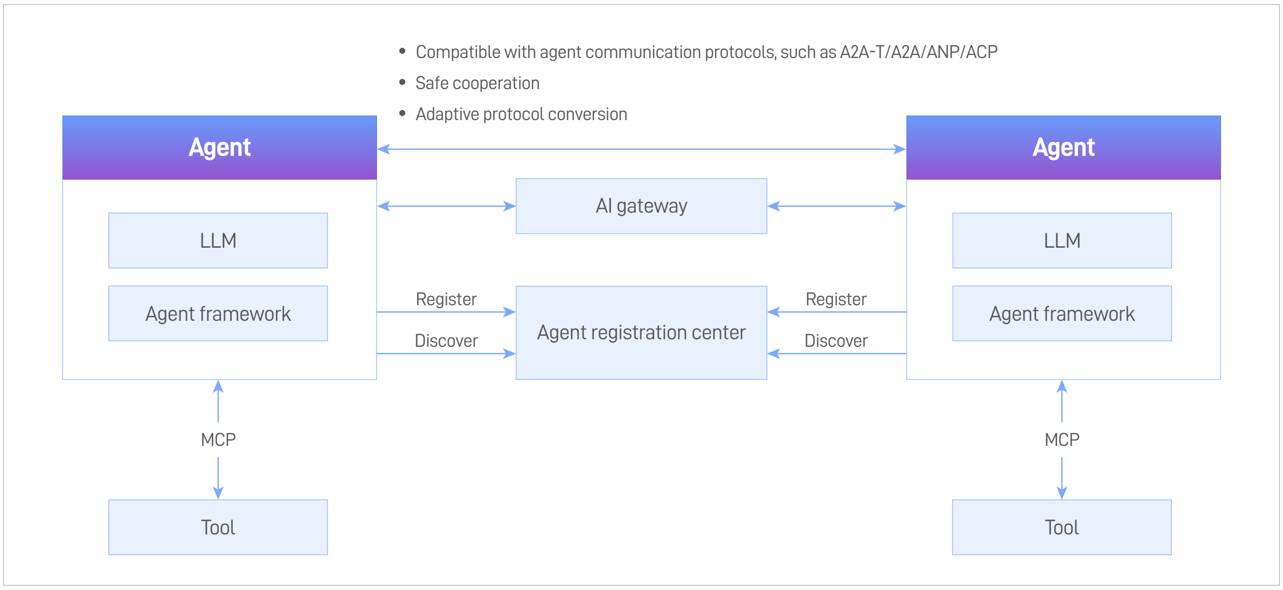

Figure 2-13 UAP Unified Proxy Protocol Architecture (Multi-agent Collaboration)

2.6.4 Digital Twin Technology

In the autonomous decision-making system driven by large models, knowledge graphs, and agents, digital twins, as the core foundation, effectively make up for the shortcomings of L4 high-level autonomous networks in decision security and simulation deduction by building high-fidelity network images. It can not only verify the security of corpora and instructions through a sandbox environment to avoid large model "illusions" and security vulnerabilities, but also provide a quantitative decision-making basis for multi-target optimization, becoming a key link to open up end-to-end autonomous capabilities.

ZTE's digital twin platform is based on simulation, decision-making, and visualization engines. By connecting the "inner closed loop of decision-making" and the "outer closed loop of virtuality and reality", it builds a complete intelligent network twin system. Relying on dynamic prediction and global optimization capabilities, the platform promotes the upgrade of network perception from "post-diagnosis" to "pre-prediction", and decision execution from "local optimization" to "global closed loop", accelerating the generational leap from passive response to active autonomous decision-making in network operation and maintenance.

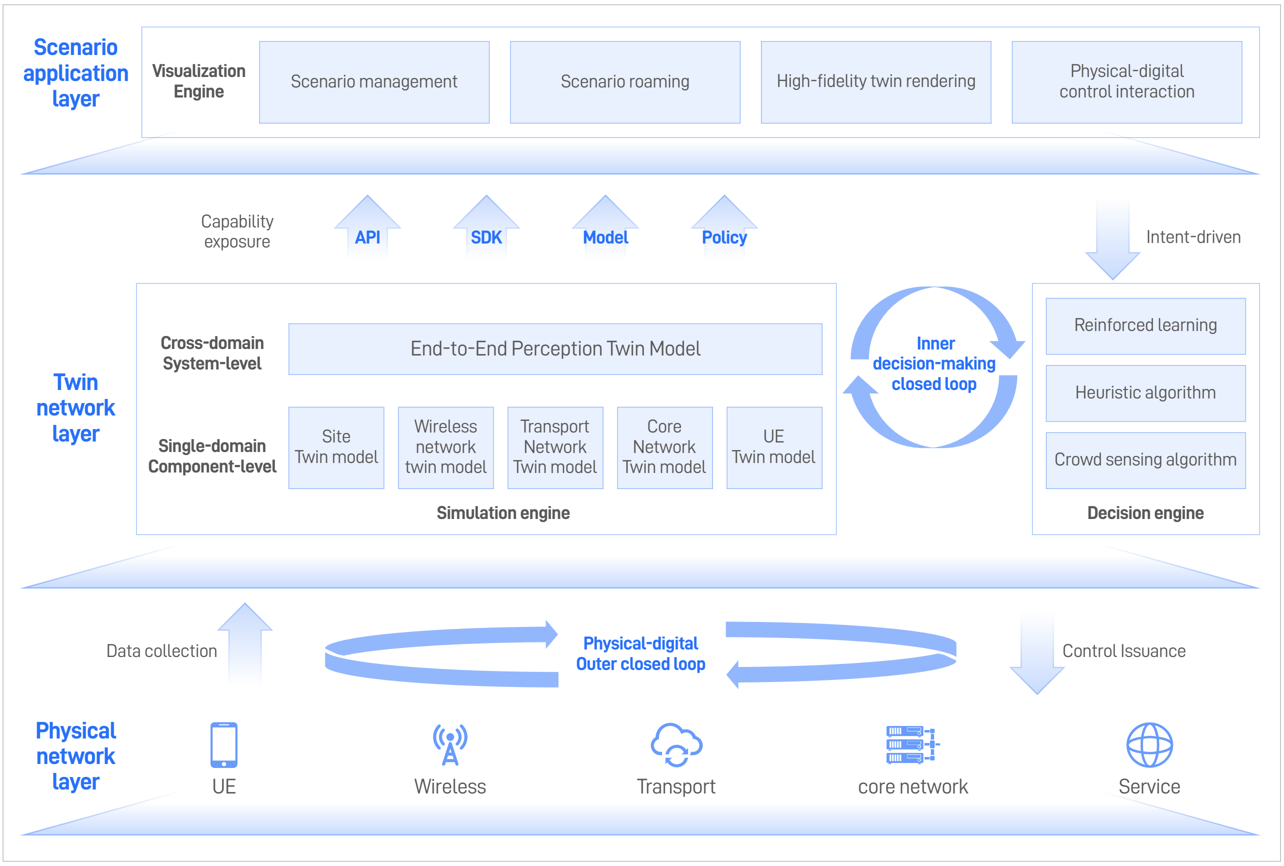

Figure 2-14 Digital Twin Network

Simulation engine:

It adopts a data and knowledge dual-driven architecture, integrating cutting-edge technology stacks such as Structured Large Model (SLM) and Graph Neural Network (GNN). With its powerful representation ability of massive heterogeneous data, SLM can accurately capture the long-distance dependence and nonlinear evolution laws of cross-domain indicators. GNN is specially designed for modeling complex communication network topologies with non-Euclidean structures, significantly improving the simulation accuracy of network topology changes and reducing simulation costs. These technologies work in conjunction with the single-domain/cross-domain twin model to build an end-to-end indicator twin system of single-domain network parameters > single-domain network KPI indicators > cross-domain perception KQI indicators to achieve accurate simulation and trend prediction of network operation status, and "sand table deduction" of optimization and adjustment plans to actively intercept invalid plans and reduce the risk of plan implementation.

Decision engine:

It integrates heuristics, mathematical programming, and deep learning algorithm libraries to adapt the optimal policy on demand according to the complexity of the scenario. It seeks the Pareto optimal solution in the simulation environment, and realizes closed-loop fine-tuning and self-evolution through the inner closed-loop mechanism of "deduction-issuance-feedback" to efficiently solve the combinatorial optimization problem under multiple constraints and multiple targets.

Visualization engine:

Relying on high-concurrency rendering and high-fidelity visual presentation technology, it deeply integrates the physical, communication, and twin worlds. The built-in Copilot intelligent assistant reshapes the interactive experience and supports temporal and spatial trajectory backtracking throughout the life cycle. Through real-time perspective and immersive roaming, it strengthens expert intervention capabilities and significantly improves collaborative operation and maintenance efficiency and control accuracy.

Taking wireless networks as an example, ZTE's digital twin platform has built a complete end-to-end indicator twin capability system from network configuration parameters to network-level performance KPIs, and then to user-level/cell-level perception KQIs, enabling the entire life cycle of network operation and maintenance. Especially in the decision execution stage, accurate mathematical modeling is performed on potential mutually exclusive targets such as network capacity, latency, reliability, and energy efficiency to generate global optimal parameter adjustment policies under complex multi-target and multi-constraint conditions, with the optimization efficiency of multi-target complex scenarios increased by 2.5 times.

2.7 Core Point 5: Process Reshaping, Breaking Through Breakpoints and Bottlenecks to Promote Work Order-Free Operation and Maintenance

The high-level evolution of autonomous networks is essentially a systematic change in the operation and maintenance paradigm from "manual drive" to "intelligent autonomy". Process reshaping is the core starting point for achieving this change. Its underlying logic lies in building an efficient operation and maintenance system with "unmanned backend processing and less manpower on the front line" through the deep coordination of capability preposition-ing and breakpoint and blockage management, and gradually realizing work order-free operation and maintenance.

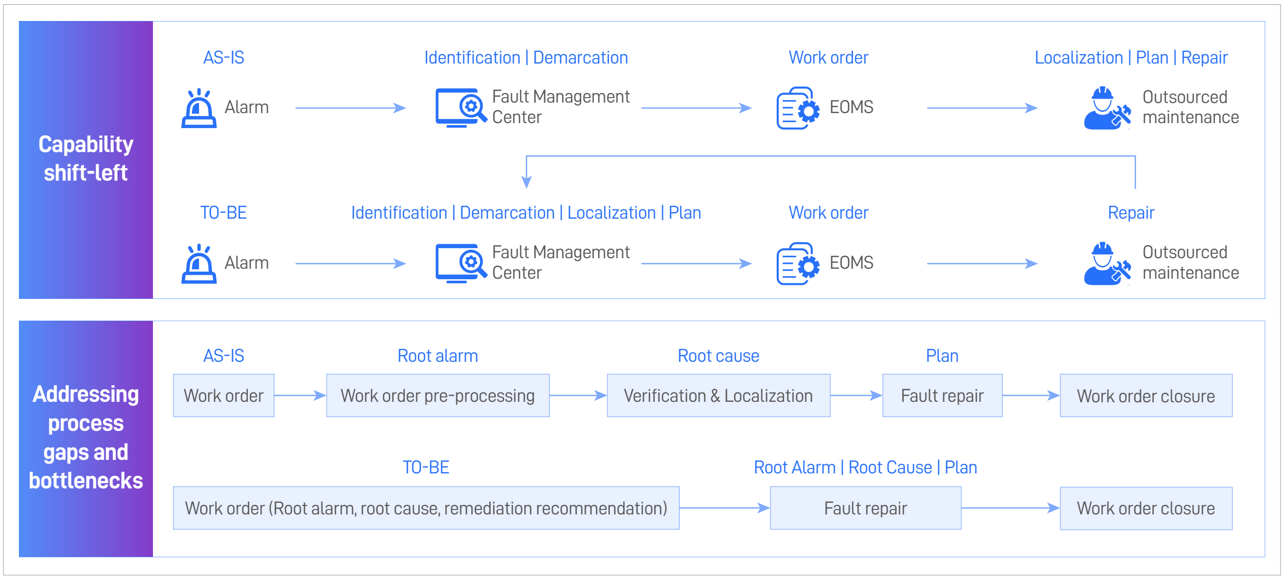

Figure 2-15 Example of End-to-End Fault Scenario Workflow Reshaping

In the traditional O&M process, core decision-making links such as cross-domain scoping, root cause localization, and solution generation are highly dependent on manual intervention, which not only leads to delayed response and inefficient coordination but also makes it difficult to support large-scale O&M requirements of the whole network due to expert experience barriers. This is not only a bottleneck for improving O&M efficiency, but also a core obstacle for autonomous networks to move towards high-level autonomy. Based on this, ZTE proposed a process reshaping idea with the goal of "unmanned backend and less manpower on the front line". Through the two-way reconstruction of "centralized decision-making capabilities in the backend and empowered execution capabilities at the front line", it broke the hierarchical barriers and data silos of traditional operation and maintenance, and gradually promoted work order-free operation and maintenance.

Based on the backend, a hierarchical intelligent decision-making system of "cross-domain delimitation + single-domain localization" is built to move the core decision-making capability forward to before the work order is dispatched. Facing the network and service, the cross-domain system automatically integrates the entire network topology, KPI, alarm, and other data to complete the impact scope definition and root alarm identification of multi-network element faults. The single domain system relies on the collaborative capabilities of large models and knowledge graphs to achieve accurate root cause localization and output standardized executable disposal guidelines. What is particularly critical is that we embed AI decision security audit links into the automated decision-making link to verify the compliance of decision logic through knowledge graphs and simulate the entire operation process based on digital twin technology to verify security. For high-risk operations such as cross-domain batch parameter modification, the system automatically triggers a secondary manual review mechanism to ensure the safety and controllability of the automated closed loop and avoid decision risks from the technical bottom.

Based on front-line work, we promote the upgrading of front-end and back-end collaborative models, and configure intelligent operation and maintenance Copilots with integrated graph model libraries, real-time consultation, and visual operation guidance for front-line operation and maintenance personnel. The decision-making results, such as root cause analysis and disposal plans output by the backend, are transformed into standardized operating procedures adapted to on-site scenarios through intelligent Copilots, which enables front-line personnel to independently complete fault handling without relying on backend experts, and greatly reduces the professional threshold and labor costs of front-line operation and maintenance.

Through the practice of end-to-end fault scenario process reshaping, we have broken through 7 breakpoints, including system switching, root alarm analysis, and front-end scheduling, and cleared 4 bottlenecks, including fault analysis, front-end and back-end interaction, and recovery verification, thereby promoting the fundamental transition of fault handling from "manual dominance, decentralized collaboration" to "intelligent drive, safety autonomy", and realizing the work order-free operation and maintenance. This not only verifies the universality and effectiveness of the idea of "unmanned backend and less manpower on the front line", but also provides a core reference for the large-scale implementation of autonomous networks in all scenarios.

2.8 Core Point 6: Business Expansion, Empowering High-Value Scenarios of Market-Network-Service Collaboration

Driven by the deepening of digital transformation and intensified market competition, the core demands of enterprise operations has shifted from scale expansion to value enhancement. Market-network-service collaboration is a key strategy to connect the front end of the market, the network base, and the service ecology. By breaking down departmental barriers, breaking through data silos, eliminating process fragmentation, and integrating fragmented resources into synergistic capabilities, it focuses on the exploration and deployment of high-value scenarios and ultimately achieves the core targets of increasing market revenue and improving operating efficiency. ZTE's autonomous network accurately anchors three core scenarios, takes technology empowerment and ecological collaboration as core levers/as key enablers/as the core approach, promotes the in-depth implementation of market-network-service collaboration, and injects momentum into the operator's market revenue increase.

Scenario 1 Precision marketing: network data driven, value mining of people-vehicle-home full-ecosystem

Taking network data as the core drive and relying on such technologies as DPI and Internet crawlers, people-vehicle-home full-ecosystem profiling of ZTE's autonomous network builds four core capabilities, including deep content analysis and "mobile network-IoT-fixed network" data integration, to accurately build a profile user ecosystem profile. Based on this profile, it can support three core applications: insight into the needs of individuals, vehicles, and home scenarios, matching customized product-service combinations, and achieving precise reach/targeted delivery at the best time through multiple channels. Ultimately, the integration of marketing and services is achieved, driving operators to transform from "selling traffic" to "meeting demands".

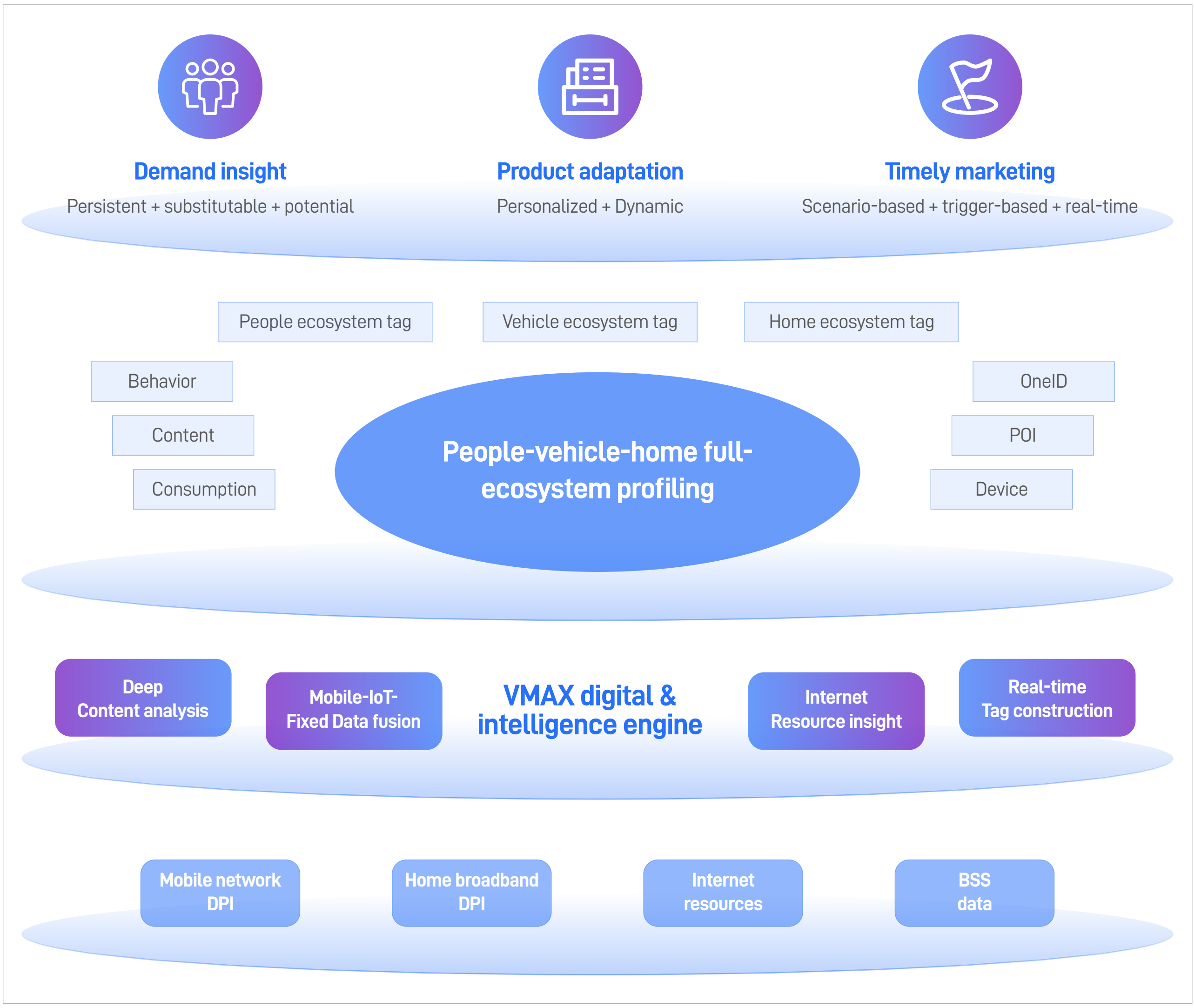

Figure 2-16 People-Vehicle-Home Full-Ecosystem Profiling

Scenario 2 Experience assurance: tiered and graded empowerment of value-based operations, transcending traditional network slicing

Focusing on users' differentiated experience demands for key services, ZTE's autonomous network builds an experience guarantee solution of "user value grading + network guarantee hierarchy". With NWDAF as the core, it integrates data from Market-Network-Service Collaboration, realizes user value grading and service demands mining through intelligent analysis, breaks through the static limitations of traditional network slicing, and realizes accurate matching of "value users-service demand-network resources". Relying on the four core capabilities of precise value classification, demand experience adaptation, resource hierarchical allocation, and cross-domain real-time guarantee, it provides differentiated experience guarantee of low latency, high reliability, and high bandwidth.

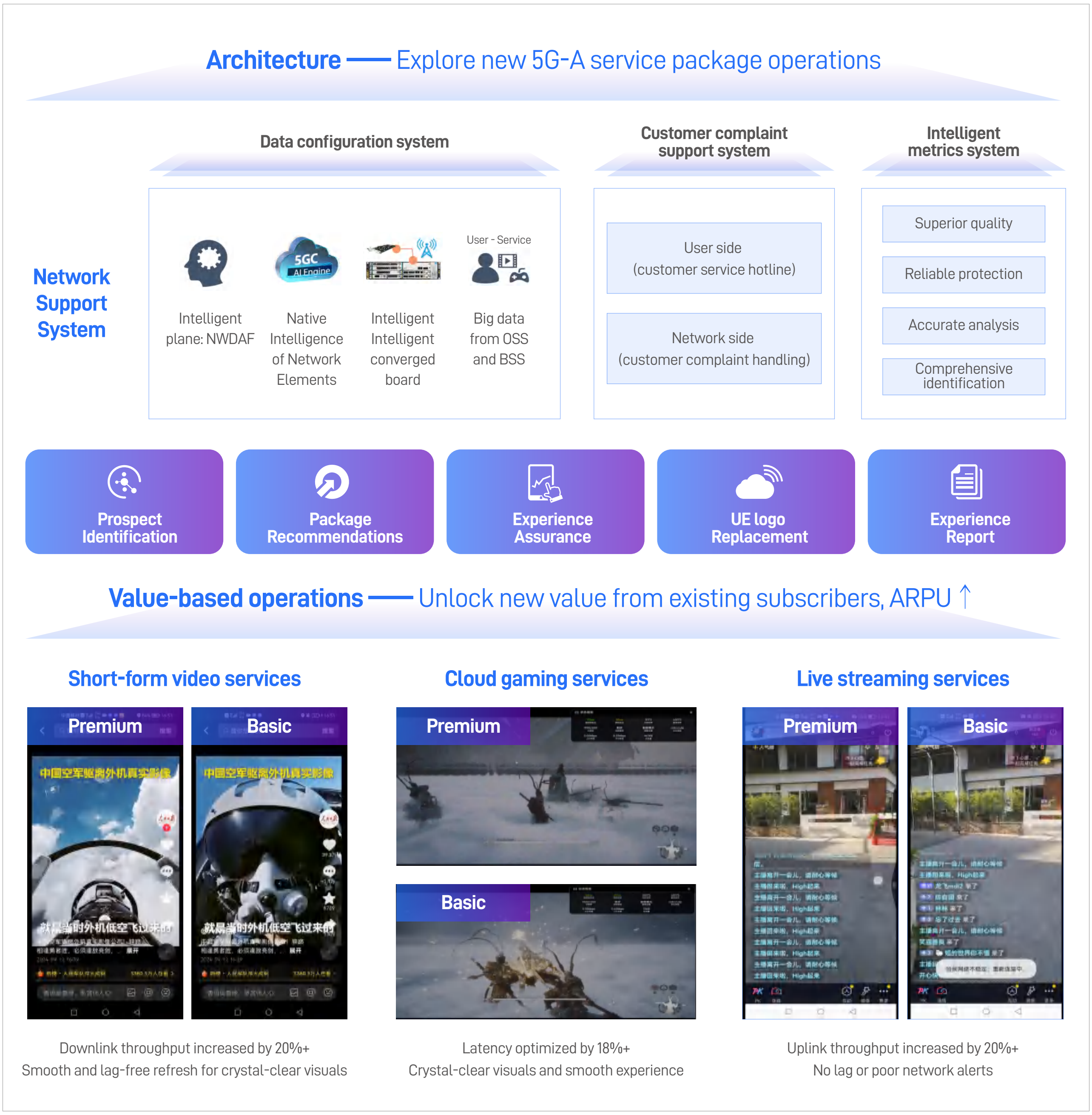

Figure 2-17 Tiered and Graded Experience Assurance

Scenario 3 Innovative operations: low-altitude economic guarantees empowering emerging industries

Relying on the autonomous network architecture, focusing on the core demands for quality assurance of low-altitude economic networks, and following the hierarchical logic of "data sources-basic capabilities-special functions-value empowerment", it integrates multi-source data such as 4/5G DPI and UAV flight records to identify, evaluate and guarantee UAV-related services, and continuously improve the network infrastructure in low-altitude scenarios, facilitating the large-scale deployment of application scenarios such as UAV logistics and emergency rescue. At the same time, it provides support for operators to expand value-added services such as low-altitude communication guarantee packages and low-altitude operation and maintenance services, forming a service closed loop of "quality assurance-industry empowerment-value realization".

Figure 2-18 Low-Altitude Economy Quality Assurance

03

L4 Practice Cases of ZTE Autonomous Networks

3.1 New Collaborative Research and Problem-solving Model: Cooperate with China Mobile to Establish the "Innovation+" Laboratory

At present, network O&M faces such pain points as highly complex data management, insufficient coverage of automated rules, and difficulty in parsing structured network data by general large models, resulting in insufficient automation capabilities for fault analysis and disposal, which restricts the evolution of high-level autonomous networks. So, it is urgent to reshape AI analysis and decision-making capabilities and accelerate the transformation of network O&M from experience-driven to data-driven.

ZTE and China Mobile joined their forces and developed a complementary partnership with autonomous scenarios as the driving force and giving full play to ZTE's deep R&D and maintenance experience advantages and China Mobile's network and data advantages. In July 2025, the "Innovation+" Autonomous Network Open Laboratory was officially established with their joint effort. The Innovation+ Laboratory aims at cutting-edge AI technology and industry standards, tackles the highly generalized, highly precise, highly reliable, and interpretable network graph model that integrates knowledge graphs with strong reasoning models, improves the autonomous analysis and decision-making capabilities of autonomous high-value scenarios, and studies the A2A-T protocol specifications for agents in the communication field to promote consensus in the industry. The two sides explored a new model of technology co-research, scenario co-creation, and ecosystem win-win, jointly invested in core R&D forces, shared scientific and technological innovation results, and jointly solved key problems in the evolution of autonomous networks.

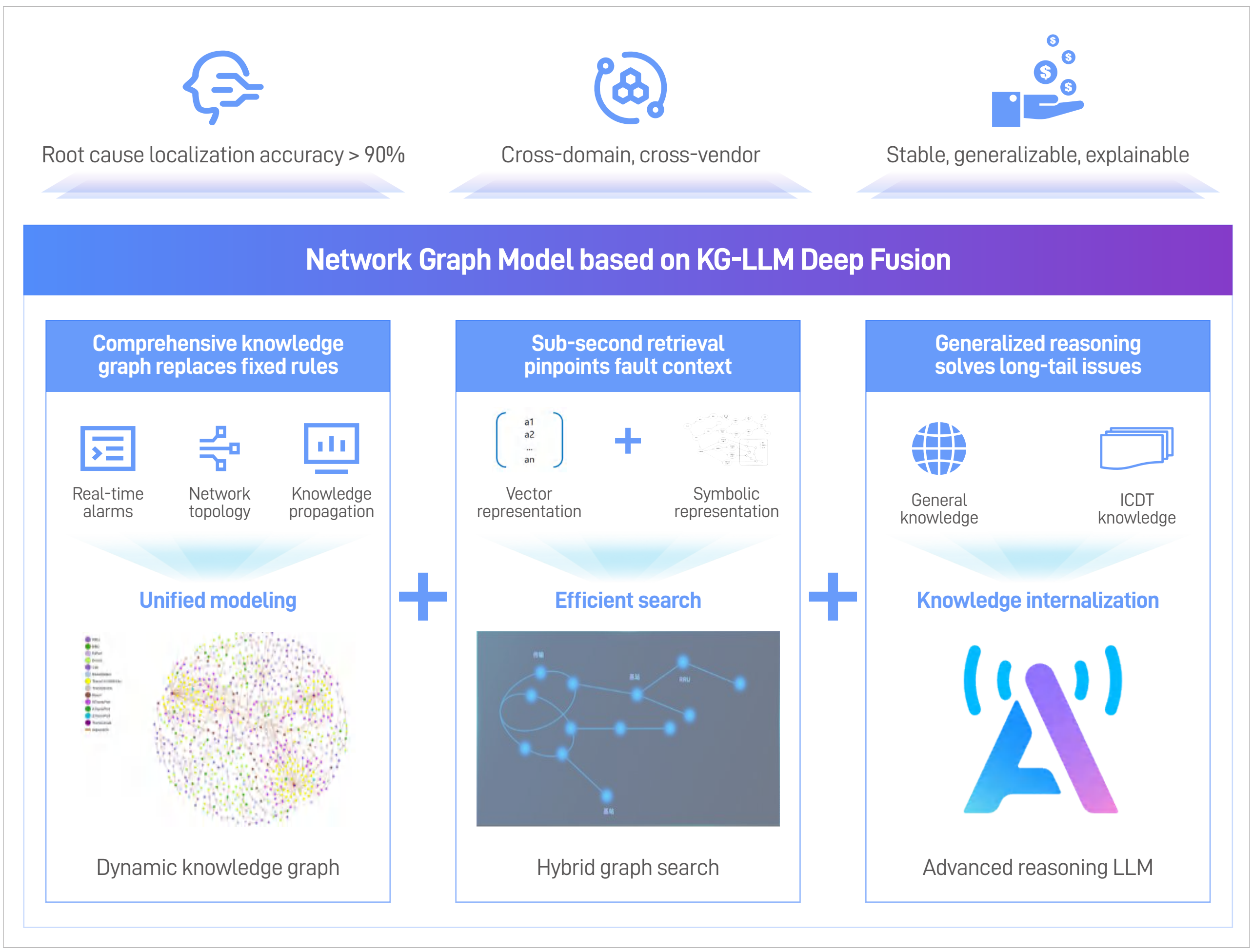

Figure 3-1 KG-LLM Fusion Reasoning

The relevant research results have been piloted in more than 30,000 base stations in Changping and Chaoyang, Beijing. It is capable of locating the root causes of faults across professions and multiple vendors. The accuracy rate of root cause localization of KCI indicators exceeds 90%. It has explored the implementation path of network graph models empowering production processes, and promoted it to Guangdong, Jiangsu, Shandong, Hunan, and other provinces for replication and application. At the same time, the two sides have jointly produced several high-quality patents, papers, and specifications, and accumulated several high-quality graph model evaluation data sets, providing valuable results for the innovation of autonomous network technology.

3.2 Intelligent Case Study: End-to-End Fault Monitoring Agent Powered by Graph Models

Background

As the cloudification of 5G networks and the integration of computing power networks accelerate, network scale and complexity continue to grow daily. Cross-domain and cross-layer faults spread rapidly and have far-reaching impacts. The current experience-driven approach to fault monitoring and handling faces significant pain points and challenges, including difficulties in locating the root causes of complex faults, reliance on manual intervention at multiple stages of operational workflows, and low efficiency in coordinating front-end and back-end fault handling.

Solution

ZTE has introduced its Nebula Communication Intelligent-Digital Employee Solution, empowering the Intelligent-Digital Employee for fault management with clearer job responsibilities and enhanced orchestration and collaboration capabilities. ZTE specifically develops Copilot-type digital intelligence agents and expert-type digital intelligence agents for fault management using high-level autonomous networks, covering all domains, including wireless networks, transport networks, core networks, network clouds, and IP networks. These solutions comprehensively reshape the end-to-end workflow of network fault identification, analysis, decision-making, handling, and evaluation. They empower maintenance professionals, NOC monitoring teams, and front/back-end outsourced maintenance personnel to significantly enhance operational efficiency while effectively reducing fault analysis and handling times. Copilot-type digital intelligence agents support O&M personnel in fault-related tasks, while expert-type agents autonomously complete the analysis closed loop for specific fault scenarios.

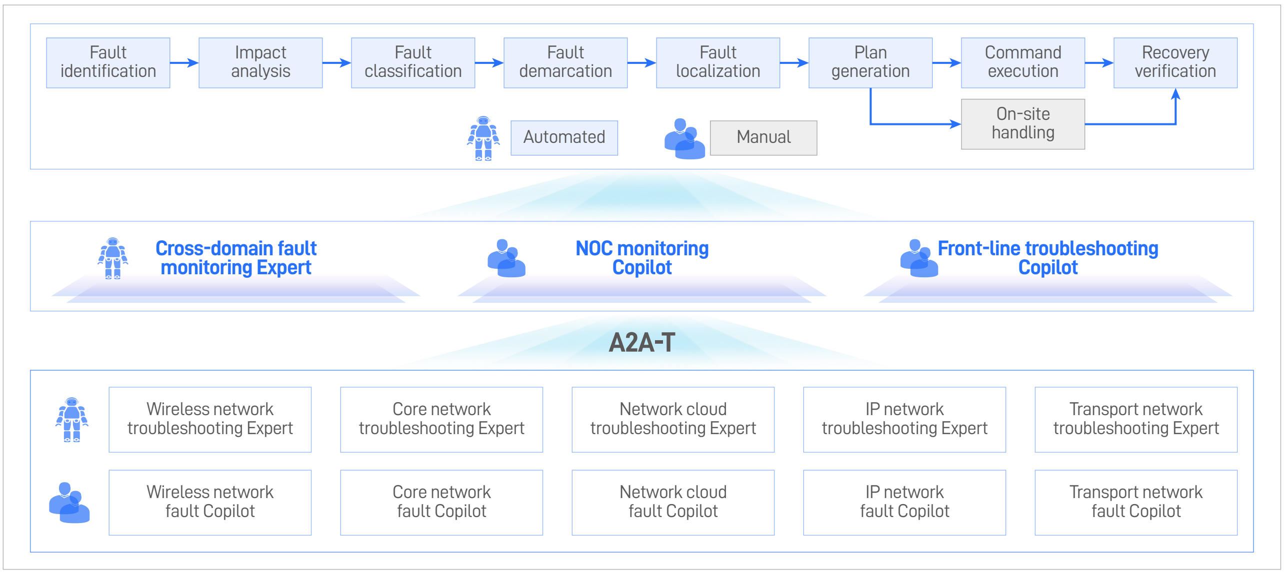

Figure 3-2 End-to-End Fault Monitoring and Handling Agent Solution

These Intelligent-Digital Employee for fault management are produced through standardized manufacturing in the Co-Sight agent factory, integrating numerous MCP services from specialized communication domains to enable seamless integration with O&M production systems. Each Intelligent-Digital Employee achieves efficient collaboration through the A2A-T agent interaction protocol, enabling closed-loop resolution of complex issues that require both cross-domain integration and intra-domain layered coordination. Building upon the general-purpose A2A foundation, A2A-T introduces security hardening for communication scenarios, adapts to the specialized contexts of the communication domain, and enhances collaborative context modeling. This delivers secure, stable, and precise collaborative capabilities.

Results

The digital intelligent agents for fault management have been deployed across over 10 sites across China and abroad, including China Mobile, China Telecom, Bangladesh GP, and Thailand AIS. These implementations empower 5 high-value professional scenarios for fault monitoring and handling, accelerating the evolution of operators' high-level autonomous networks. ZTE provides comprehensive network O&M management services for GP Bangladesh, a key subsidiary of the Telenor Group. By deploying troubleshooting experts and troubleshooting Copilots, ZTE enhances end-to-end monitoring and handling efficiency, achieving a 12.5% reduction in fault MTTR (KEI), a 70% improvement in human-machine collaboration efficiency, and a 3% reduction in outstanding fault work orders.

In wireless network fault monitoring and handling scenarios, the troubleshooting experts integrated into the workflow for wireless fault root cause localization and analysis, along with the troubleshooting Copilot empowering frontline maintenance personnel with self-service troubleshooting capabilities, achieved a 21.34% reduction in the fault MTTR (KEI). The solution was honored with the Pioneer Award of the 2025 TM Forum Copenhagen Innovation Hub.

In core network fault monitoring and handling scenarios, the digital intelligence agents for fault management have significantly improved fault handling efficiency. The fault MTTR (KEI) was reduced by over 20%, and the average fault diagnosis time per work order was shortened to under 10 minutes. The solution was included in the China Academy of Information and Communications Technology's case studies for LLM applications in the telecommunications industry.

In network cloud fault monitoring and handling scenarios, the digital intelligence agents for fault management achieved a 75% improvement in root cause diagnosis efficiency, with a root cause diagnosis accuracy (KCI) of 96.3%, and an automatic alarm diagnosis coverage rate of 92.63%.

In transport network fault monitoring and handling scenarios, the digital intelligence agents for fault management reduced the optical cable fault MTTR (KEI) from 5.3 hours to 3.2 hours, and shortened the recovery time for 100 service lines from 2 hours to 20 minutes. The solution was honored with the first award of the ICT China (2025) Cases.

In cross-domain fault monitoring scenarios, the digital intelligence agents for fault management empower cross-domain and cross-vendor intelligent fault identification, demarcation, localization, and analysis across IP networks, transport networks, power and environment networks, wireless networks, and network clouds. The complex fault demarcation and localization accuracy (KCI) increased to 91%, while the fault MTTR (KEI) decreased by 11%. The solution was honored with the 2025 AIIA Outstanding Case Award for the Agent Category.

3.3 Intelligent Case Study: SPN Network Change Expert Powered by LLM Technology

Background

As cybersecurity and service quality requirements continue to rise, traditional network change approaches are increasingly unable to meet the demands for security, intelligence, and efficiency. Network changes constitute major operational tasks characterized by extended operation times and significant impact. The overall workflow encompasses the following stages: plan generation and analysis, pre-change verification, change execution, post-change validation, and post-change monitoring. Due to its large scale, diverse services, and complex change scenarios, the SPN network requires extensive involvement from O&M personnel during each network change. The average operation time is 13.9 hours, involving numerous manual steps that are prone to errors. These errors may cause network service interruptions and lead to major failures. For example, operations on the core ring may cause widespread base station service interruptions, resulting in significant economic losses.

Solution

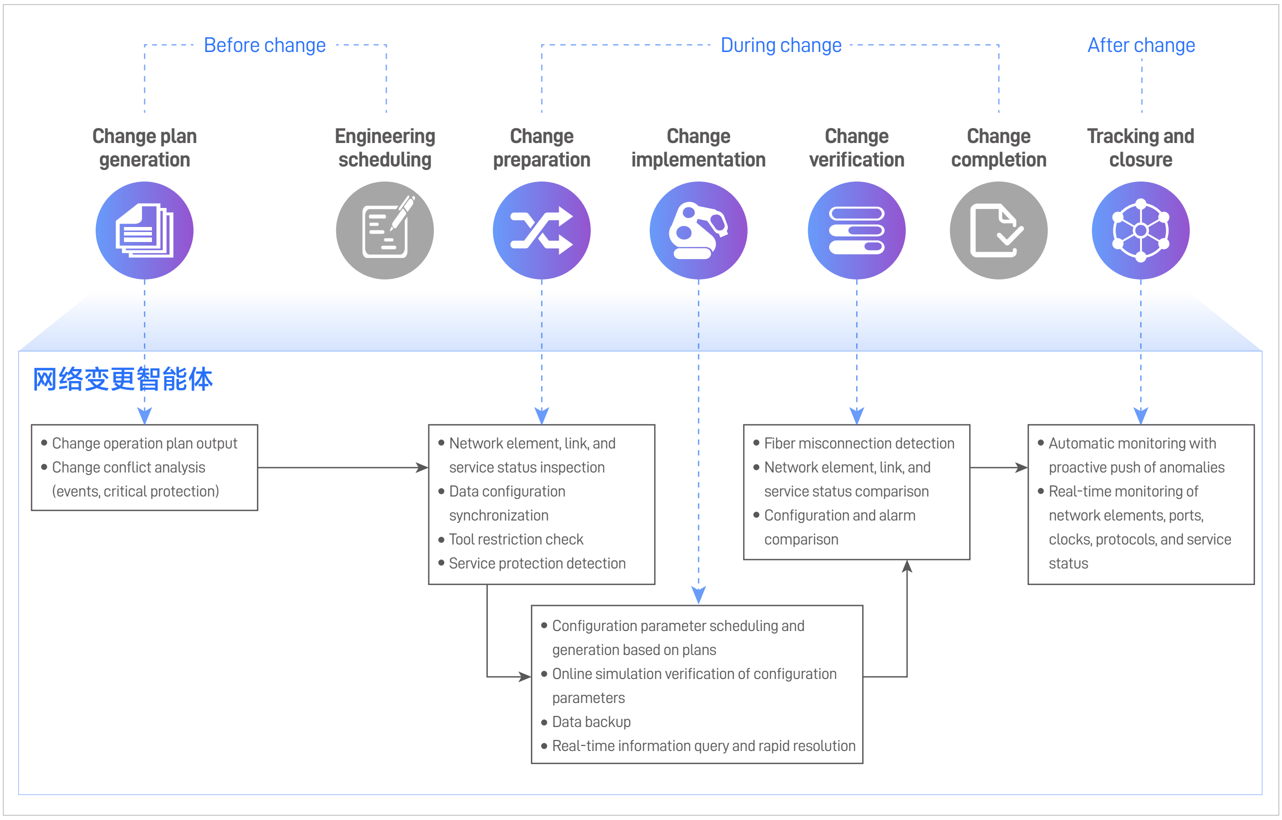

Figure 3-3 SPN Network Change Expert Solution

ZTE launches the industry's first SPN Network Change Expert Solution to address the pain points of network change. Firstly, building upon the existing natural language-driven sequential network change workflow, the solution further enhances control over the workflow. It achieves a fully automated closed-loop O&M encompassing change plan generation, change preparation, change execution, change verification, and change monitoring. This shifts the approach from "human-led system guidance" to "system-guided human involvement" for network changes, minimizing manual intervention throughout the entire workflow while significantly improving change operation efficiency. Secondly, for SPN network change scenarios, the solution leverages diverse network O&M data (including industry standards, technical specifications, product manuals, expert experience summaries, and engineering case studies) to undergo targeted training and fine-tuning. This enables the solution to automatically generate network change plans and the configuration parameters required for change operations. The solution flexibly adapts to various network change scenarios while mitigating the uncertainty risks associated with manual design and configuration parameter generation. Thirdly, leveraging a digital twin foundation, the solution conducts a comprehensive real-time simulation of the agent-generated plans and configuration parameters. It outputs reports on three network metrics and three service evaluations, enabling a holistic assessment of the plan's impact on both network and services before change execution. Finally, after confirming that the change task is accurately completed, the agent activates its automated monitoring function. It performs comprehensive real-time monitoring of the affected network areas, proactively pushing alarms for any network or service anomalies to guide O&M personnel in rapid resolution.

The solution employs multiple innovative and leading-edge technologies, enhancing network change efficiency while ensuring network and service security.

Results

The solution reduced the average operation time in the SPN network change (KEI) from 13.9 hours to under 3 hours, achieving an overall efficiency improvement of over 78%. It also completely prevents human errors caused by network change operations. The solution has received multiple honors, including the second award of the ICT China (2025) Cases, highlighting its leadership position and social impact within the industry.

3.4 Intelligent Case Study: Home Broadband Service Complaint Handling Expert Enabling Efficient Complaint Handling

Background

With the rapid growth of FTTX home broadband users, their focus has shifted from bandwidth to overall experience. Issues such as slow internet speeds often trigger user complaints. Traditional analysis methods relying on PON network alarms and KPIs frequently fail to precisely demarcate and locate the fault root causes. These methods lack end-to-end service-level monitoring and the analytical tools for KQIs. The challenges that operators urgently seek to address include proactively responding to complaints, identifying the root causes of home broadband complaints based on end-to-end service flow quality data, precisely demarcating and locating the fault root causes, swiftly resolving complaints, and ultimately enhancing user satisfaction with home broadband services.

Solution

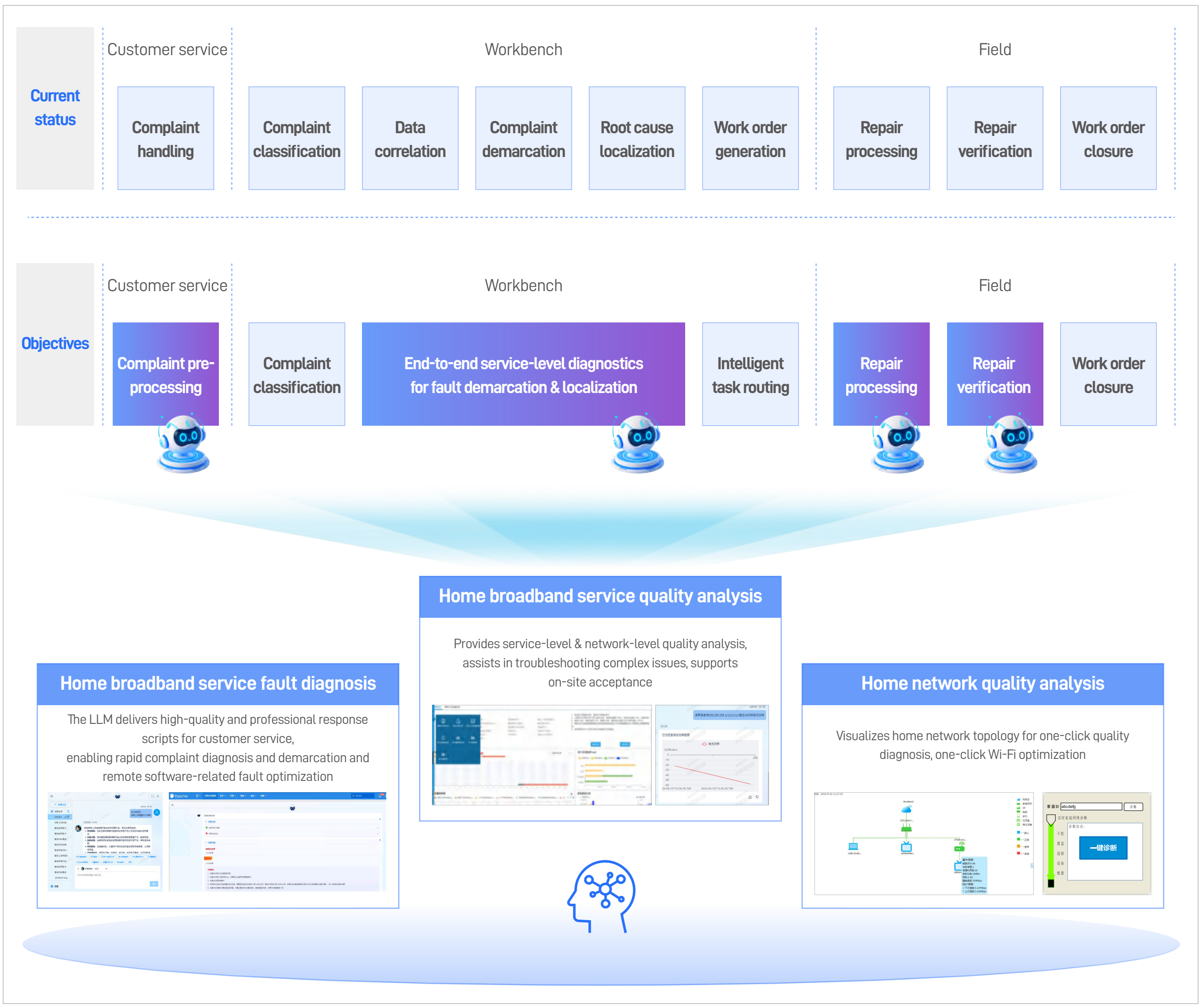

Figure 3-4 Home Broadband Service Complaint Handling Expert Solution

ZTE launches the industry's first Home Broadband Service Complaint Handling Expert Solution to address the pain points of home broadband complaint handling. The solution integrates with operators' customer service centers, NOC dashboards, and field technical Copilots to support complaint handling, as well as fault root cause demarcation and localization. It enables intelligent task routing while empowering frontline technicians to enhance on-site efficiency and effectiveness.

The solution offers support across the three stages of complaint handling:

1. Intake Stage: The Home Broadband Service Complaint Handling Expert offers end-to-end pre-diagnostic capabilities to identify issues like service degradation or fiber breaks. The agent precisely demarcates and locates the fault root causes, enabling intelligent task routing for complaints. For software-related faults, the agent enables one-click remote fixes to reduce the volume of fault work orders.

2. Analysis Stage: The Home Broadband Service Complaint Handling Expert enables precise root cause demarcation and localization of access network segment faults, further aiding fault assignment and rapid repair. For complex issues, it supports NOC center personnel by enabling service-level and network-level quality traceability alongside efficient and expedited analysis capabilities.

3. Maintenance Stage: The Home Broadband Service Complaint Handling Expert/Copilot supports field technical Copilots with home network fault analysis capabilities and Wi-Fi analysis self-optimization functions; it assists field technicians in rapid repair and verification.

Results

The solution enables end-to-end fault diagnosis for home broadband user complaints, empowering the integrated task routing center to perform pre-diagnosis before task routing and identify whether service degradation resides in one of the following five segments: home network, access network, OLT-BRAS segment, BRAS-content source segment, or content source. For service degradation, the root cause demarcation and localization accuracy (KCI) exceeds 90%. Furthermore, the solution provides root cause analysis and maintenance recommendations for the identified service degradation, enabling intelligent task routing and processing through the work order system, significantly enhancing operational efficiency in complaint handling. By precisely demarcating and locating reported faults, the solution enables intelligent task routing, significantly reducing redundant work order transfers. Analysis results provide maintenance recommendations, aiding field technicians in rapid repairs, reducing the on-site MTTR (KEI) by approximately 20%. The solution enhances operators' home broadband user satisfaction and product competitiveness. The solution was honored with the "Outstanding FTTH Service" award at the 2024 Network X exhibition.

3.5 Intelligent Case Study: Manageable and Controllable Wireless Network Optimization Based on Digital Twins

Background

Wireless network optimization has long been plagued by numerous optimization tasks and extended optimization cycles, making it one of the areas where operators concentrate significant resource investments. For network performance optimization scenarios, the core objectives are primarily reflected in two aspects: first, comprehensively enhance users' service experience through proactive optimization of network quality; second, reduce reliance on mid-tier maintenance and outsourced maintenance resources through intelligent means, effectively controlling operational expenditures. To achieve the above objectives, it is essential to enhance the automation and intelligence levels in critical areas such as issue identification, root cause analysis, localization, and policy generation. This involves two key approaches: first, expanding the coverage of automated rules and AI capabilities; second, continuously improving the accuracy of identification and handling. These efforts will systematically advance the closure rate of automated wireless network optimization.

Solution

Digital twin technology empowers wireless network optimization by enabling real-time and dynamic identification of service degradation affecting UEs and precise root cause localization (physical-to-digital mirroring). Leveraging live network data, a high-fidelity digital twin model for wireless KPIs is constructed (digital-to-physical prediction). Taking the reduction of wireless network issues as the primary optimization objective, and subject to the constraints that no other indicators degrade and no new problematic cells are introduced, the system performs multi-objective and multi-cell optimization under multiple KPIs and network constraints to identify the optimal solution. The digital twin simulation for the performance of hundreds of parameter plans ultimately yielded the optimal solution satisfying all optimization objectives and the anticipated multi-objective optimization gains (digital-to-physical optimization).

After implementing the twin optimization plan, the digital twin platform conducts a post-optimization assessment. This evaluates whether issues have been resolved while also comparing digital twin metrics against actual on-site metrics post-issuance to assess the digital twin fidelity. Data from both pre- and post-optimization phases is collected to construct new training samples for model self-iteration, ensuring the twin environment continuously maintains a 1:1 faithful representation of the live network (digital-physical symbiosis).

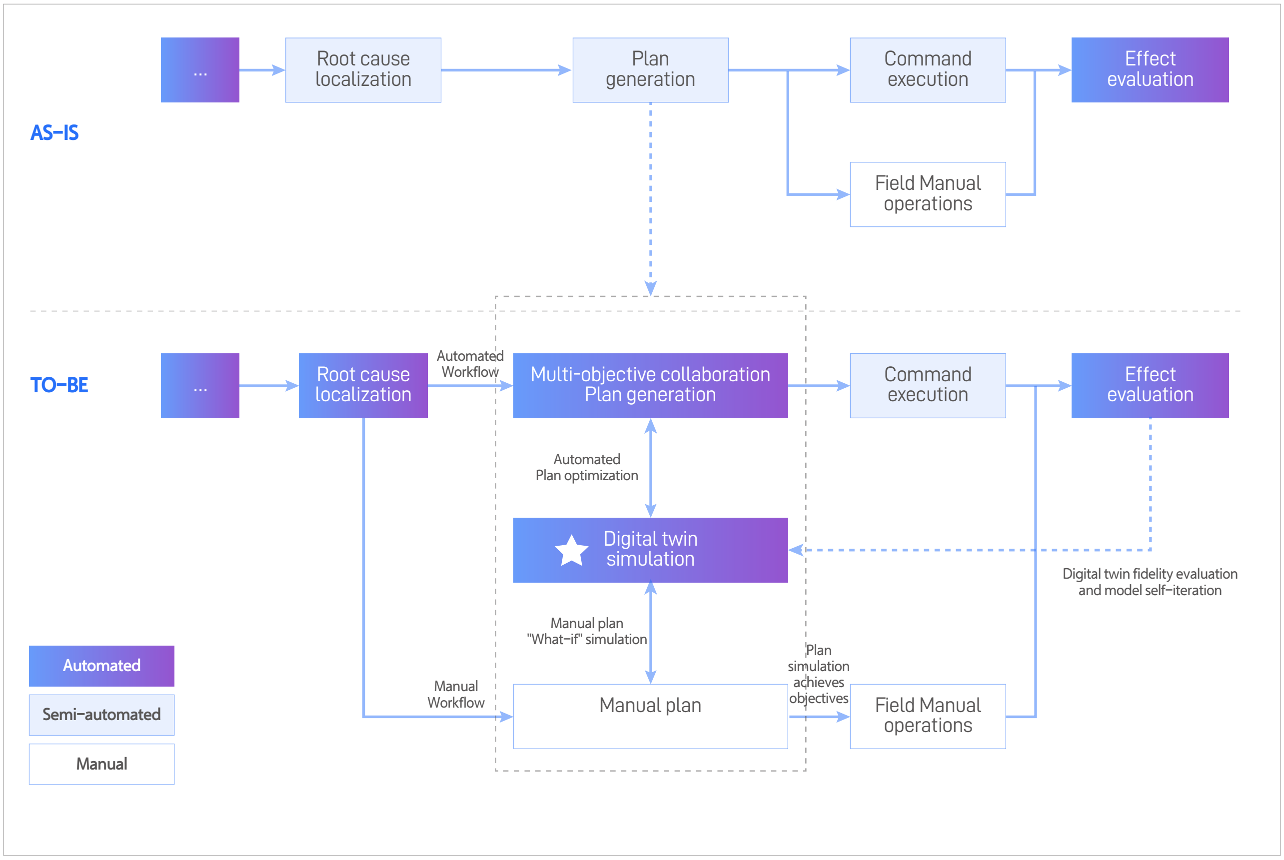

Figure 3-5 Manageable and Controllable Network Optimization Solution Based on Digital Twins

Results

During a small-scale pilot verification involving 800 base stations at China Mobile Fujian, the plan generation time (KCI) was reduced from an average of 3.81 hours to 0.5 hours. The proportion of work orders requiring secondary optimization decreased from 31.45% to 12.9%. Wireless network performance optimization has seen significant improvements in both time and efficiency during the core plan generation phase.Deployment and promotion will be expanded in the follow-up phase.

3.6 Intelligent Case Study: Quantitative Simulation Evaluation of Resource Pool Switchover Based on Digital Twins

Background

In DC-level disaster recovery switchover scenarios, pre-recovery resource assessment relies heavily on human judgment and experience, consuming hours of time. This approach is prone to uncontrolled signaling storms due to insufficient personnel expertise and skills, potentially triggering secondary failures. Compounded by the lack of visibility during the switchover process, risks become unmanageable. Faced with signaling impact peaks caused by large-scale and multi-service concurrent switchover, existing systems lack forward-looking simulation and quantitative decision-making capabilities.

Solution

ZTE has innovatively developed a "Core Network Digital Twin Simulation System", driving a fundamental shift in disaster recovery switchover from "experience-driven" to "data-driven" approaches. The system collects real-time live network topology and signaling performance data through the data layer, constructs high-fidelity digital twin simulation models at the model instance layer, and integrates topology visualization, automatic flow control parameter verification, and millisecond-level signaling impact simulation capabilities at the twin application layer. This enables quantitative simulation assessment of resource pool switchover, supporting stable, agile, and secure core network services. Furthermore, the system enables the opening of digital twin service capabilities to third parties through APIs.

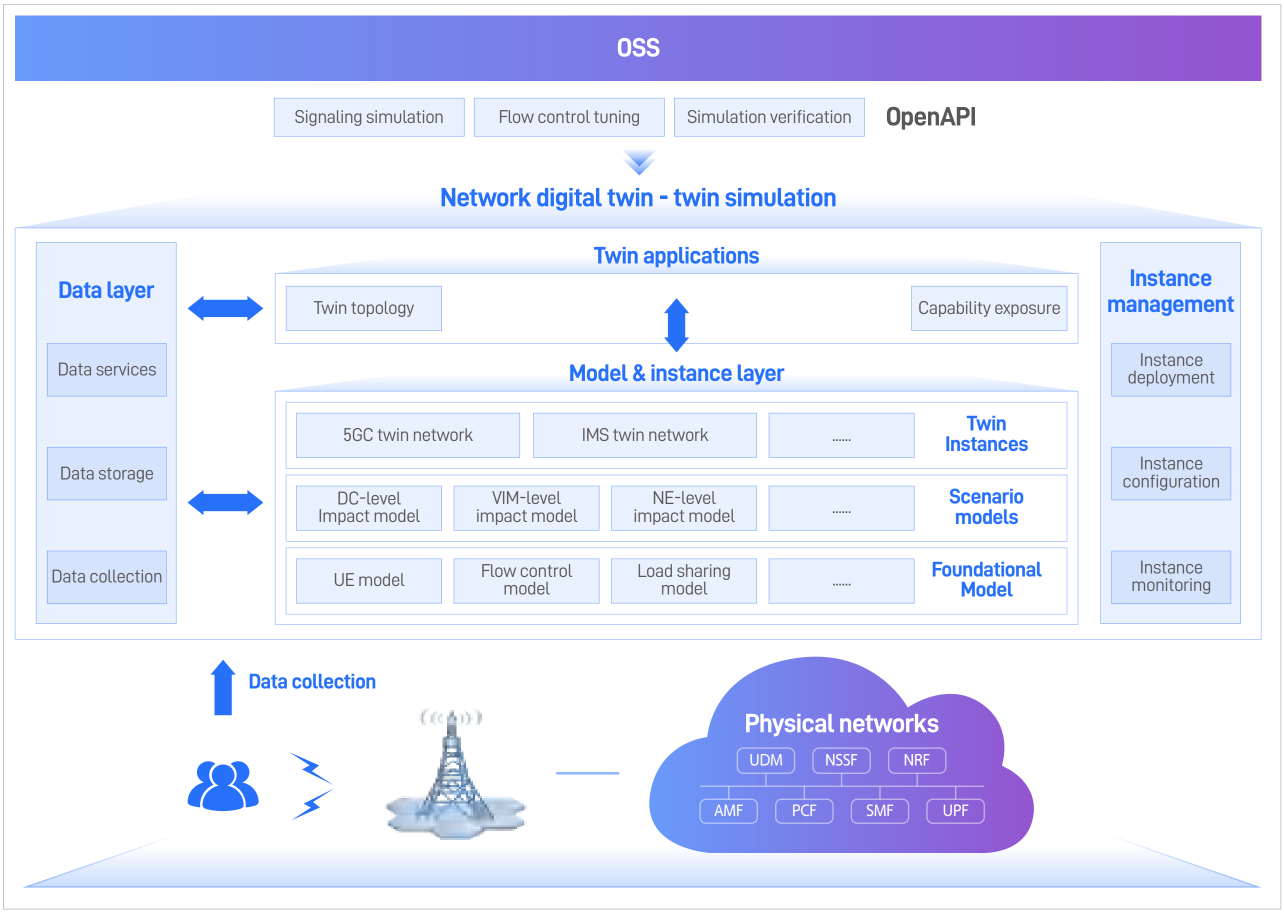

Figure 3-6 Core Network Digital Twin Simulation System

Results

The evaluation time (KCI) was reduced from an average of 4 hours to under 18 minutes. The prediction accuracy of signaling impacts (KCI) through simulation reached 92%, and the coverage rate for switchover risk identification (KCI) exceeded 95%. The solution was honored with the second award of the ICT China (2025) Cases, establishing itself as a benchmark practice for intelligent disaster recovery O&M in core networks.

3.7 Value-Based Operation Case Study: NWDAF-Driven Tiered and Graded Experience Assurance Solution Empowers 5G-A Monetization

Background

With the widespread adoption of 5G technology, users' demands for network service experiences continue to rise, and expectations for personalized service assurance grow increasingly stringent. By implementing a tiered and graded experience assurance solution for the core network, we achieve precise allocation of network resources to safeguard user services and enhance the overall user experience. With the commercial deployment of the tiered and graded experience assurance solution for core networks across multiple scenarios and the continuous enhancement of AI capabilities, the application of the innovative solution will significantly expand the commercial value of 5G networks and accelerate the digital transformation of society.

Solution

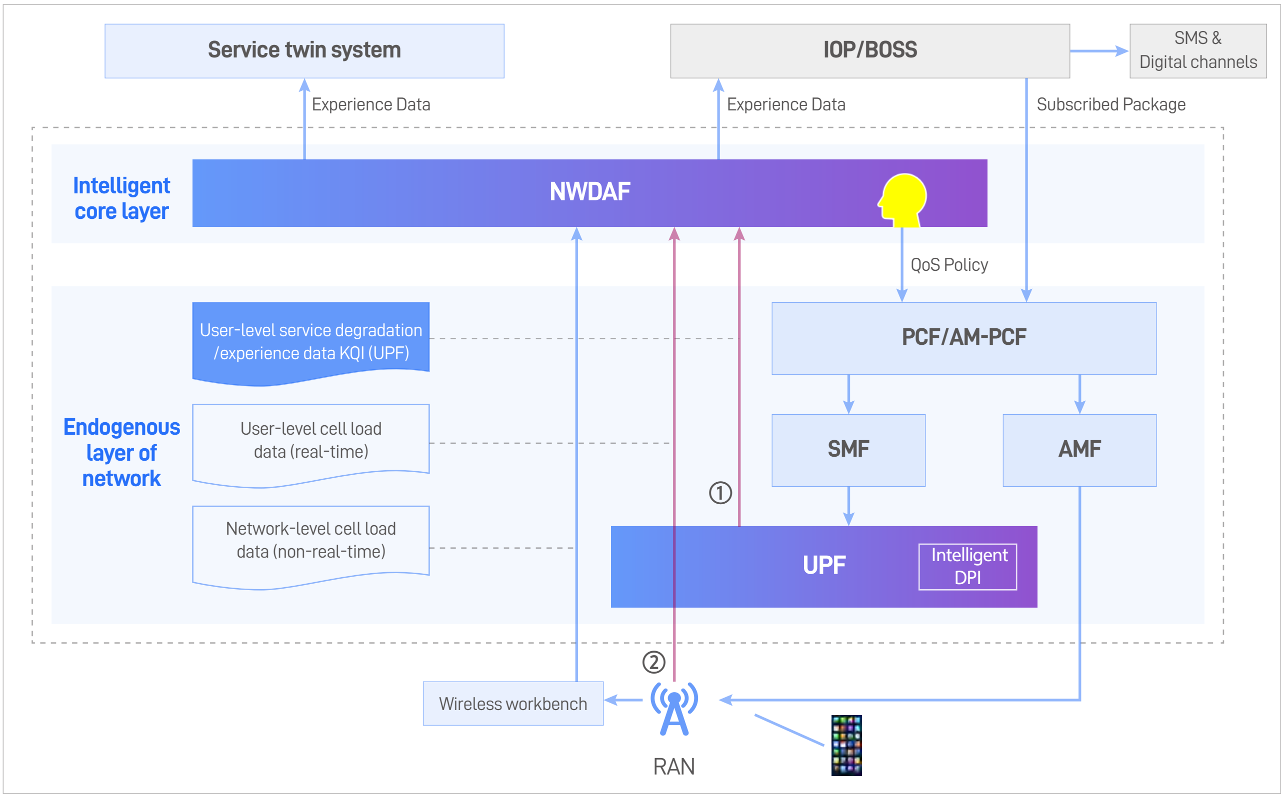

Figure 3-7 NWDAF-Driven Tiered and Graded Experience Assurance Solution

The tiered and graded experience assurance solution rapidly identifies service degradation for individual users and promptly triggers targeted assurance actions to enhance user experience. To address the challenges of precise service degradation handling and experience-aware assurance, as well as the visibility of assurance effectiveness, the solution is implemented through an end-to-end integrated approach of "UPF - NWDAF/PCF - Service Twin". The UPF handles granular data collection and identification, the NWDAF dynamically generates QoS assurance policies, the PCF enables flexible customization and display of UE logos, and the Service Twin system visualizes assurance outcomes and simulates policies. Through coordinated collaboration among these modules, the solution achieves tiered and graded network assurance while enhancing the user's perceived service experience.

Results

Scale of Tiered and Graded Experience Assurance: By the end of 2025, the service had been enabled for tens of millions of VIP users across China.

Enhancement in Tiered and Graded Experience Assurance Effectiveness: Assured users triggered assurance actions 3.65 times per day on average; average assurance duration per user (KEI) was 9.57 minutes; throughput for prioritized TikTok traffic (KCI) increased by 35.6%.

ZTE and China Mobile Guangdong have jointly launched an innovative initiative. The solution was honored with the 2024 SDN/NFV/Network AI Outstanding Case Award and the first award of the ICT China (2025) Cases.

04

Industry Collaboration Initiative for Evolution of L4 Highly Autonomous Networks

I. New journey for evolution of highly autonomous networks

2025 marks the inaugural year of L4 AN commercial deployment, with high-level autonomous networks accelerating their evolution from single-domain closed-loop systems to cross-domain collaboration, from operational optimization to value creation, and from technology validation to large-scale commercial deployment. Over the next 3-5 years, the evolution will exhibit three core characteristics:

1. Capability Deepening: From "Single-Point Intelligence" to "Cross-Domain Collective Intelligence". Leveraging the Agentic AI architecture and A2A-T protocol, the enhancement for closed-loop capabilities of individual agents enables cross-domain and cross-layer intelligent orchestration. Simultaneously, the deep fusion of LLMs, digital twins, and knowledge graphs achieves end-to-end trustworthy intelligence spanning "perception - analysis - decision-making - execution", empowering precise capabilities in complex scenarios.

2. Value Enhancement: From "Cost Reduction and Efficiency Improvement" to "Ecosystem-Driven Value Co-creation". By driving cross-functional collaboration among municipal governments, network operators, and vertical industries, 5G-A and Industrial Internet use cases enable large-scale monetization of network capabilities. The network has become the core infrastructure for AI-native applications, unlocking a hundreds-of-billions-yuan market. Furthermore, the KBI-KEI-KCI framework aligns with industry-specific metrics, establishing a closed-loop system for value realization.

3. Architectural Innovation: From "Full-stack AI" to "AI-native Networking". The Agentic AI architecture enables deep integration between networks and AI, empowering systems with multi-framework compatibility and heterogeneous computing resource scheduling capabilities. Driven by dual engines of AI-native intelligence and software twins, it accelerates the development paradigm transformation. Trusted AI technologies ensure end-to-end lifecycle security, laying a robust foundation for large-scale deployment.

II. Industry collaboration initiative: jointly building a new ecosystem for highly autonomous networks

1. Collaborative Standards Alignment: Drive collaboration among global standards organizations to accelerate the development of critical standards; leverage the TM Forum Autonomous Network Level Assessment and Validation (ANLAV) framework to advance commercial standardization, establish a "generic + specialized" standards architecture, and build an effectiveness evaluation system encompassing KBI-KEI-KCI.

2. Joint Technology Innovation: Collaborate across industry, academia, research, and application stakeholders to build innovation platforms centered on core technologies, accelerate breakthroughs, open interfaces and verification environments, and establish patent-sharing mechanisms to reduce costs and improve efficiency.

3. Collaborative Scenario Development: Operators and vertical industries join forces to advance the large-scale implementation of high-value scenarios, distilling replicable best practices; explore the NaaS business model to achieve mutual value creation.

4. Ecosystem win-win: Build a platform for the entire industry chain to engage SMEs; open software and hardware resources to empower evaluation capabilities; strengthen international cooperation to share practical experience.

III. Conclusion

Autonomous networks are reshaping the architecture and business models of the telecommunications industry, becoming the core infrastructure of the digital economy. Only by prioritizing value, focusing on technology, and pursuing collaboration can we accelerate the implementation of autonomous networks. Therefore, ZTE looks forward to joining hands with the entire industry chain, embracing an open and inclusive mindset and taking pragmatic and innovative actions to jointly build a new ecosystem for highly autonomous networks. May we contribute greater strength to the development of Digital China and the advancement of the global communications industry!

Acronyms

Acronym	Full Name in English
5G-A	5G-Advanced
GenAI	Generative AI
Agent	Agent
AN	Autonomous Networks
ANL	Autonomous Networks Level
CCSA	China Communications Standards Association
CHB	Customer, Home, Business
CO-DA	Copilot Data
CO-Tuning	Copilot Tuning
DTW	Digital Transformation World
ENI	Experiential Networked Intelligence
ETSI	European Telecommunications Standards Institute
GNN	Graph Neural Network
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
NGMN	Next Generation Mobile Network
NOC	Network Operation Center
ODN	Optical Distribution Network
OMC	Operation and Maintenance Center
OPEX	Operating Expense
OTN	Optical Transport Network
PON	Passive Optical Network
SLM	Structured Large Model
SOC	Service Operation Center
SPN	Slicing Packet Network

References

[1] TM Forum. Autonomous Networks Implementation Guide. June 2025

[2] China Mobile. China Mobile Autonomous Networks White Paper (2025)

[3] China Telecom. China Telecom Autonomous Cloud-Network Operations White Paper 4.0 (2024)

[4] China Unicom. China Unicom Autonomous Networks White Paper (2025)

[5] Xi Jinping. Continuously Strengthen, Optimize, and Expand China's Digital Economy. Qiushi Journal, February 2022

[6] Ericsson. Elevating 5G with Differentiated Connectivity. December 2024

To Enable Connectivity and Trust Everywhere

ZTE Corporation