

Expert Views

Internet of Agents: Redefining
the Terminal-Network
Collaboration Paradigm

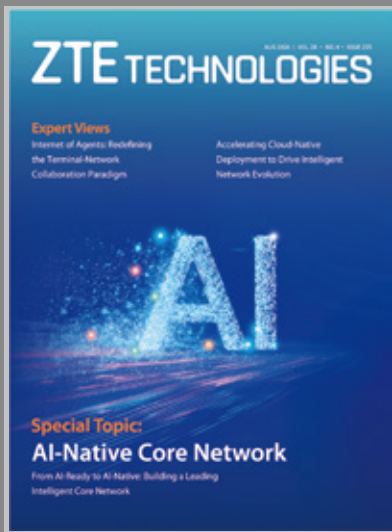
Accelerating Cloud-Native
Deployment to Drive Intelligent
Network Evolution

Special Topic:

AI-Native Core Network

From AI-Ready to AI-Native: Building a Leading
Intelligent Core Network





ZTE TECHNOLOGIES

AUG 2026 | VOL. 28 • NO. 4 • ISSUE 225

Advisory Committee

Director: Liu Jian

Deputy Directors: Fang Hui, Sun Fangping, Peng Aiguang, Zhang Wanchun

Advisors: Bai Gang, Chen Xinyun, Dong Weijie, Hu Junjie, Hu Lihua, Kan Jie, Li Qiang, Li Xiaotong, Tang Xue, Wang Quan, Zheng Peng

Editorial Board

Director: Lin Xiaodong

Deputy Director: Lu Dan

Members: Deng Zhifeng, Dai Yanbin, Guan Kai, Huang Xinming, Liang Dapeng, Lin Xiaodong, Ma Xiaosong, Shi Jun, Wang Weibin, Xiao Wei, Yang Zhaojiang, Yu Fanghong, Zhao Jianchao

Sponsor: ZTE Corporation

Published by ZTE Technologies Editorial Office

General Editor: Lin Xiaodong

Deputy General Editor: Lu Dan

Editor-in-Chief: Liu Yang

Circulation Manager: Wang Pingping

Editorial Office Address: No. 55, Hi-tech Road South, Shenzhen, P.R. China

Circulation Office Address: 12F Kaixuan Building, 329 Jinzhai Road, Hefei, P.R. China

Website: www.zte.com.cn/global/about/publications

Email: magazine@zte.com.cn

Statement: This magazine is a free publication for you. If you do not want to receive it in the future, you can send the "TD unsubscribe" mail to magazine@zte.com.cn.

We will not send you this magazine again after receiving your email. Thank you for your support.

Toward AI-Native Transformation, Led by the Core Network



Chen Xinyu

Vice President of ZTE and General Manager of Computing Power & Core Network Products

The industrialization of AI has entered an explosive phase. The proliferation of AI agents and on-device AI is reshaping telecom networks. Networks are evolving from "bit pipes" to "token-centric services" and from providing connectivity to delivering intelligent digital services. Currently, operators face the dual challenges of stagnant revenue growth and the urgent need for AI transformation. To lead in the AI-native era, they must achieve coordinated breakthroughs across architecture, business, and ecosystems.

As the hub of mobile networks, the core network plays a key role in perception, decision-making, optimization, and operations. It is evolving from a "network brain" to an "intelligent neural hub." As the critical link connecting AI technologies with various vertical industries, the core network supports the transition toward AI-native operations and enables operators to rebuild competitive advantages through AI.

The AI-native core network vision is built around AI agents, reconstructing network capabilities in perception, decision-making, and execution. From intent understanding to autonomous

evolution, it transforms the network into a value engine featuring experience-based monetization, self-healing operations, and scalable ecosystems to meet AI-driven communication needs. Across four dimensions—business, connectivity, O&M, and beyond connectivity—it enables four major upgrades: enhancing interaction, perception, and service entry points to reshape voice and messaging services; evolving toward personalized connectivity, realizing experience-based operations and stimulating traffic growth; establishing agent-driven O&M with enhanced interaction, optimization, and predictive capabilities; and expanding data, security, and computing services to open new growth avenues for operators.

At this critical juncture, as the industry embraces AI-native transformation, we present our insights and practices in AI-native core networks. We aim to foster industry-wide innovation, advance intelligent service ecosystems, and accelerate the evolution of networks from connectivity platforms to intelligent hubs, unlocking value of AI-network convergence.

CONTENTS



Expert Views

- 02 Internet of Agents: Redefining the Terminal-Network Collaboration Paradigm

By Jin Youxing

- 06 Accelerating Cloud-Native Deployment to Drive Intelligent Network Evolution

By Yang Lin

Success Stories

- 35 ZTE Launches Complaint Analysis Agent to Redefine the International Roaming Service Experience

By Zhang Maopeng

- 38 Network Operations Agent (NOA) V1.0: Enabling a New Paradigm of Intelligent Connectivity

By Yuan Xincheng, Wei Beibei, Chen Xiaofeng

- 40 Partnering with China Mobile to Establish the "Co-Innovation+" Autonomous Network Open Lab

By Zhou Bo

Special Topic: AI-Native Core Network

- 10 From AI-Ready to AI-Native: Building a Leading Intelligent Core Network

By Wang Weibin, Lu Guanghui

- 15 AI-Native 6G Core Network: Shaping a New Paradigm of Intelligent Services

By Zhou Jianfeng, Zheng Xingming

- 19 Redefining Experience Operations to Unlock Network Monetization

By Cao Yilin

- 22 AI Calling: A Strategic Elevation from Voice Portal to AI Life Portal

By Lu Jiwen

- 24 "Car-Home Connect": A Pan-Domain Private Network for People, Vehicles, and Homes

By Liu Xiliang, Si Hairui

- 28 Core Network O&M Agents: Enabling Continuous Evolution of L4 Autonomous Networks

By He Wei

- 31 Integrated Compute-Network-Intelligence Infrastructure for 6G

By Zhu Kun

- 33 AI+ Signaling Network: Efficient and Reliable Signaling Hub

By Zheng Guobin

Internet of Agents: Redefining the Terminal-Network Collaboration Paradigm



Jin Youxing

Chief Engineer of CCN Product Planning, ZTE

As artificial intelligence evolves from perceptual intelligence to actionable intelligence, AI agents are emerging as a key bridge between the digital and physical worlds. A new-generation intelligent ecosystem is taking shape, with agent-enabled devices as access points and the Internet of Agents (IoA) as the connectivity framework. This transformation not only redefines human-computer interaction but also reshapes device forms and network architectures, driving the evolution of information networks toward intelligent collaborative networks.

The Evolution of Agent Terminals: From Passive Response to Proactive Service

Powered by AI agents, agent terminals possess closed-loop capabilities encompassing perception, decision-making, action, and learning, enabling them to evolve from passive execution devices into cognitive intelligent partners that proactively understand user intentions and autonomously complete tasks, thereby achieving the shift from "passive response" to "proactive service" and delivering more intelligent, personalized, and efficient user experiences.

Embodied intelligence, as a typical agent terminal,

not only possesses the core capabilities of AI agent—namely perception, decision-making, action, learning—but also interacts with the real environment through a physical body, such as humanoid robots, quadruped robot dogs, and wheeled-arm robots, thereby achieving intelligence that can both "think" and "act" in the physical world.

Agent terminals have the following characteristics:

- **Multimodal perception and interaction:** Integrate various sensors, such as vision, hearing, and touch, to achieve comprehensive perception of the environment.
- **Collaborative reasoning:** Equipped with lightweight models, supporting local reasoning, edge-network collaborative reasoning, etc.
- **Autonomous decision-making and execution:** Based on user input and perceptual data, analyze and make decisions using AI models, and execute corresponding operations.
- **Low power consumption:** Adopt various innovative technologies to improve computational efficiency and reduce energy consumption.
- **Digital identity authentication:** Terminal authentication methods have evolved from SIM card-based authentication to digital identity authentication/authorization.

Agent terminals impose higher requirements on networks, including large uplink capacity, low latency, high reliability, massive connectivity, and heterogeneous collaboration.

- **Large uplink capacity:** Besides local processing, agent terminals need to continuously upload multimodal sensing data, including videos, images, audio, point clouds, and tactile streams, for cloud-based model inference, world model simulation, and teleoperation.
- **Low latency:** Physical interactions of embodied intelligence demand millisecond-level response; otherwise, unstable movements or safety incidents may occur.
- **High reliability:** The physical world allows very little margin for error. A single communication interruption may cause embodied intelligent devices to fall, collide or fail to complete tasks.
- **Massive connectivity and heterogeneous collaboration:** Terminal-based agents usually work with other agents, edge service nodes, and cloud platforms to accomplish complex tasks.

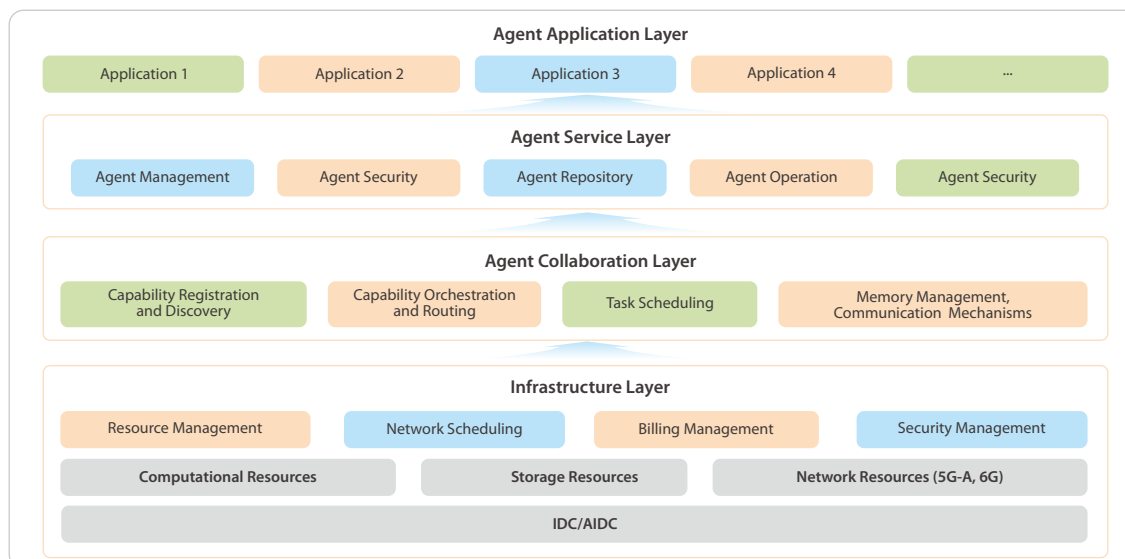
Restructuring Network Infrastructure: From Connectivity Pipelines to Intelligent Hubs

The advancement of agent terminals will drive the evolution of traditional networks and the Internet toward the Internet of Agents (IoA). The IoA is a digital ecosystem composed of autonomous agents

that can discover, communicate and cooperate with one another. With autonomous agents serving as the core interaction entities, the IoA adopts unified protocols and interfaces to enable autonomous discovery, efficient interaction, and collaboration among agents, between humans and agents, and between agents and tools. This marks a paradigm shift of the Internet from "connecting information" to "connecting intelligence" (see Fig. 1).

The Internet of Agents features the following attributes:

- **New connectivity entities:** Humans and agents stand on an equal footing, jointly forming a human-agent collaboration ecosystem. As digital workers, agents perform functions such as perception, autonomous planning, and task execution.
- **New interaction modes:** Traditional CLI/GUI interactions are replaced by multimodal natural interactions, including voice, gestures, facial expressions, and eye contact, with support for intent understanding and context awareness.
- **New collaboration models:** Collaboration shifts from human-led operations to autonomous agent collaboration. This generates substantial east-west traffic, transforming networks from pure data pipelines into intelligent collaboration hubs.
- **New resource orchestration:** Resources such as data, networks, computing power, and tools are highly abstracted to enable on-demand dynamic scheduling and realize capability-as-a-service.
- **New governance mechanisms:** Agents achieve



◀ Fig. 1 Diagram of the Internet of Agents architecture.

decentralized collaboration via consensus mechanisms, negotiation, or federated learning. Multiple agents can spontaneously form task alliances to fulfill complex goals without centralized scheduling, greatly enhancing system robustness and scalability.

With core capabilities including semantic addressing, zero-trust security, and behavioral contracts, the Internet of Agents is not merely an upgrade of the existing Internet. Instead, it treats agents as native participants and builds a new infrastructure that integrates humans, machines, and intelligence.

Edge-Network Collaboration: Reshaping the Terminal-Network Collaboration Paradigm

The rise of agent terminals and the Internet of Agents is profoundly reshaping the terminal-network collaboration paradigm, with impacts far exceeding those of the traditional mobile Internet.

Reversed Traffic Model

Traffic patterns have shifted from downlink-dominated to uplink-dominated. While traditional mobile phones are primarily used for video downloads, agent terminals need to continuously upload multimodal sensing data, including 4K/8K videos, point clouds, tactile streams, and IMU data. The uplink bandwidth requirement of a single device can exceed 200 Mbps.

Stringent Requirements for Latency and Reliability

When embodied intelligence interacts with the physical world, even millisecond-level latency may lead to falls or collisions. The development of the Internet of Agents is driving networks to evolve from best-effort delivery to deterministic services.

Expanded Connectivity Capabilities

Collaboration scenarios are expanding from human-network interaction to human-machine, machine-machine, and machine-tool cooperation. This generates massive volumes of highly concurrent east-west traffic, reshaping the traditional dominant north-south traffic pattern.

Evolved Network Functions

Networks are evolving from pure transmission pipelines to intelligent collaboration hubs. Beyond transmitting and securing communication data for agents, networks now participate in intelligent decision-making. Typical applications include conveying key intents via semantic communication, integrating network-native intelligence with services such as computing-network convergence, understanding upper-layer task semantics to dynamically allocate differentiated resources, and enabling agent addressing and cross-domain interconnection.

New Security and Privacy Requirements

Networks need to adopt a zero-trust architecture and support end-to-end encryption, intelligent behavior tracing, and safety circuit breaker mechanisms as fundamental capabilities. This ensures that agents operate reliably and compliantly in open environments.

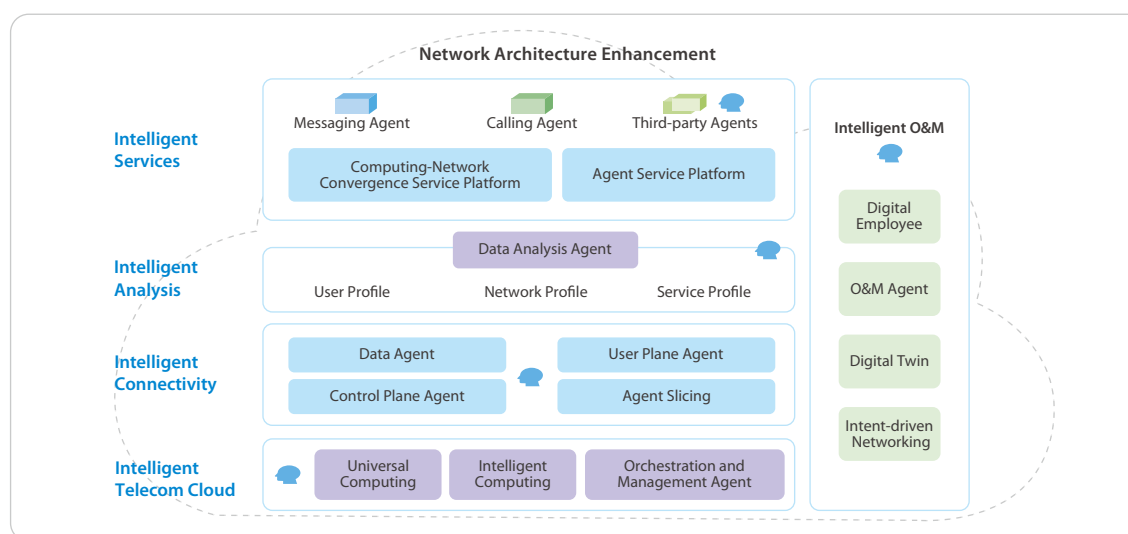
Agent terminals and the Internet of Agents are transforming networks from information highways into intelligent neural hubs. Capable of intent understanding, security assurance, and cross-domain collaboration, networks have become the cornerstone for the large-scale deployment of the Internet of Agents.

Internet of Agents: Driving Network Innovation and High-Quality Growth for Operators

The rise of the Internet of Agents is driving networks to shift from "connecting people" to "connecting agents and serving the physical world", which requires operators to comprehensively restructure their network architectures, operational models, and service systems.

Restructuring of Network Value

Networks are evolving from connection-centric services to multi-dimensional converged services, involving changes in service targets and revenue models. The service scope is expanding from individual users to enterprise agent clusters. Service-level agreements (SLAs) now incorporate new metrics including task success rate, deterministic latency, reliability, and safety-triggered circuit-breaker response. Meanwhile, revenue models are transforming from traffic-based charging to capability-based charging, covering semantic



◀ Fig. 2 Schematic diagram of the upgraded network architecture.

communication, security services, innovative networking capabilities, agent-enabled services, and AI content generation services.

Restructuring of Network Architecture

Build a new AI-native infrastructure integrating cloud, network, and intelligent computing capabilities: deploy deterministic networks by introducing enhanced URLLC, time-sensitive networking (TSN), and integrated sensing and communication technologies to support high-value agent services and embodied intelligence services; develop edge intelligence by deploying MEC and AI inference nodes in industrial parks and factories to enable lightweight model deployment and edge-terminal-cloud collaborative services; establish a multilevel computing network to form a “terminal-edge-cloud” computing hierarchy, dynamically schedule computing and network resources based on task complexity, and deliver optimized overall performance.

Restructuring of Network Services

Operators are evolving from pure pipeline providers to converged service providers, with network architectures adapted to new service positioning (see Fig. 2).

- **Dedicated network slices for agents:** Customize network slices for industrial agents, deliver differentiated ToB slice services, and optimize agent communication services.
- **Agent service platform:** Provide one-stop

services including registration, discovery, authentication, billing, and monitoring to support full lifecycle management of agents. On this basis, develop industry-specific solutions, launch self-developed agents, and introduce third-party service providers to deliver Agent-as-a-Service (AaaS). Build a developer ecosystem by opening network capability APIs to attract independent software vendors (ISVs) to develop agent applications, forming a dual-engine model driven by network capabilities and applications.

- **Converged computing, network and intelligence services:** Offer integrated computing, network, and AI services to enable in-depth collaboration and deliver value-added services beyond basic connectivity.

The Internet of Agents is not merely a technical upgrade, but a paradigm shift that transforms the physical world into a collaborative, evolvable, and trustworthy intelligent ecosystem. It elevates human-machine interaction from tool use to symbiotic collaboration, reshapes the industrial landscape, and charts a path for future network infrastructure. Through its AIR MAX solution, ZTE has turned the integration of AI agents and networks from a concept into practical deployment. By improving energy, operational, and investment efficiency, AIR MAX establishes a new paradigm for AI-era mobile networks and empowers operators in business transformation and value creation. **ZTE TECHNOLOGIES**

Accelerating Cloud-Native Deployment to Drive Intelligent Network Evolution



Yang Lin

Chief Engineer of CCN Product Planning, ZTE

Over the past decade, with the introduction of NFV technology, cloud computing has become a key force driving the transformation of telecom network architecture. In particular, the inherent advantages of cloud-native technologies—such as lightweight design, component decoupling, and continuous integration and continuous delivery (CI/CD)—have made cloud nativeness not only the preferred approach for new 5G SA core network deployments, but also a key pathway for network modernization. According to Omdia's 2025 research report, the share of containerized network functions (CNFs) in global commercial networks has reached 24%, and is projected to exceed 35% by 2030, surpassing both PNFs and VNFs.

Meanwhile, as AI transformation has become a global trend, the telecom industry is actively exploring how to utilize AI. According to a GSMA Intelligence report, AI has been the top-priority technology for telecom operators for two consecutive years since 2024. However, since cloud has become the foundation of telecom networks, failing to clarify the relationship between cloud-native development and intelligence evolution may not only delay this evolution, but also waste current investments.

Intelligence: The Current Focus and Consensus of Network Evolution

Intelligence has become a key driving force behind the ongoing evolution of multiple global telecom standards, including those developed by 3GPP, TMF, and ETSI.

Since the first 5G SA standard (Release 15), 3GPP has regarded intelligence as a critical requirement and one of the core technology directions for subsequent enhanced releases. As shown in Fig. 1, the architecture has evolved from an initial centralized NWDAF-based framework for network data collection, analysis, and feedback, to a distributed and optimized architecture that decouples data collection, storage, AI training, and AI inference functions. The latest architecture further enables cross-domain collaboration between network and management data, as well as federated learning across networks to support more diverse service scenarios and flexible deployment requirements. In 6G, AI-native architecture which supports coordinated operation of multiple AI systems, and native AI services, which extend connectivity and empower intelligent systems, are becoming explicit design targets for 3GPP.

Since 2019, TM Forum has positioned autonomous networks (AN) as a top strategic priority for driving

the digital transformation of the communications industry. By defining the six autonomous network levels (ANL) evaluation system from L0 to L5, TM Forum has established a quantifiable and certifiable global standard for network autonomy, which has become a evolution target for both global operators' networks and vendors' products.

In 2024, the AN White Paper 6.0 explicitly introduced GenAI (represented by the foundational large model TelcoGPT) and agentic AI to achieve L4 and higher autonomy. TelcoGPT is used not only for critical network data analysis to help network operation engineers quickly localize faults and provide handling suggestions, but also for potential fault pre-alerts based on multi-dimensional analysis of data including network hardware alarms, network performance indicators, and network traffic. Meanwhile, two types of agentic applications, namely Copilots for operation roles and agents for operation scenarios, have become key to enhancing autonomy across all layers.

In addition to the clear roadmap for standards and technology evolution, AI has recently become a focus for operators, driven largely by business pressures, including the long-standing need for cost reduction and the pursuit of new AI-driven revenue streams such as experience monetization. All mainstream operators have set achieving AN L4

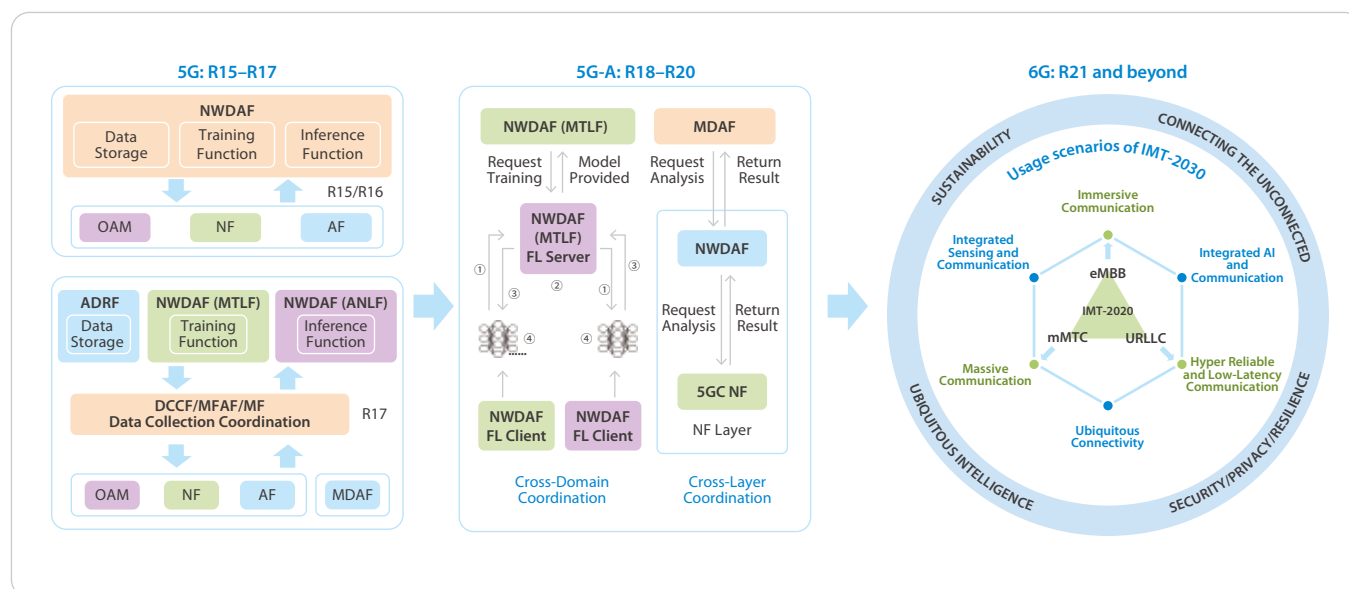
before 2030 as a strategic goal.

Facing the strong demand for intelligence, network vendors have not only integrated more AI capabilities into existing products, but also made AI a key focus of future product planning. However, building a cloud platform that provides end-to-end AI services—from data preparation and model training to AI inference—is critical to enabling this evolution.

Kubernetes, the De Facto Standard for AI Workload Orchestration and Deployment

Alongside enterprise digital transformation and the adoption of CNCF open-source projects as de facto cloud-native standards, Kubernetes (K8s) and other CNCF ecosystem technologies have become core choices for enterprises to build modern application architectures. According to industry data, 98% of organizations have adopted cloud-native technologies, with 82% of container users using K8s for deployment in the production environment. Now K8s is gradually becoming the default choice for deploying new AI software for the following reasons.

First, K8s has become the dominant management platform for AI accelerators, including GPUs, TPUs, and NPUs from different vendors. It provides powerful resource management and optimization



▲ Fig. 1. Intelligence evolution of 3GPP and 6G requirements.

capabilities for AI applications, such as efficiently allocating and managing GPU/TPU/NPU resources through its Scheduler and Extender mechanisms. It also enables AI applications to automatically adjust resource allocation based on business needs. To better support the management and integration of AI-specific hardware, K8s offers comprehensive mechanisms, including drivers, device plugins, and GPU monitoring. Major GPU/TPU/NPU vendors prioritize integration with K8s as the first step toward entering the AI software ecosystem.

Second, K8s' scalability enables the large-scale deployment of AI applications, allowing developers to extend Kubernetes by defining custom resource definitions (CRDs) and Operators to manage complex AI workflows. Meanwhile, K8s-based observability tools, such as OpenLLMetry, provide comprehensive monitoring and logging capabilities for AI applications, helping developers quickly identify and resolve performance bottlenecks and ensure the stability and reliability of AI inference services. For AI workloads including large models and AI applications, new CNCF open-source projects such as Fluid, Kubeflow, and KServe can cover the full lifecycle from data preprocessing and model training to inference workload deployment and LLM integration.

Third, popular high-performance large model

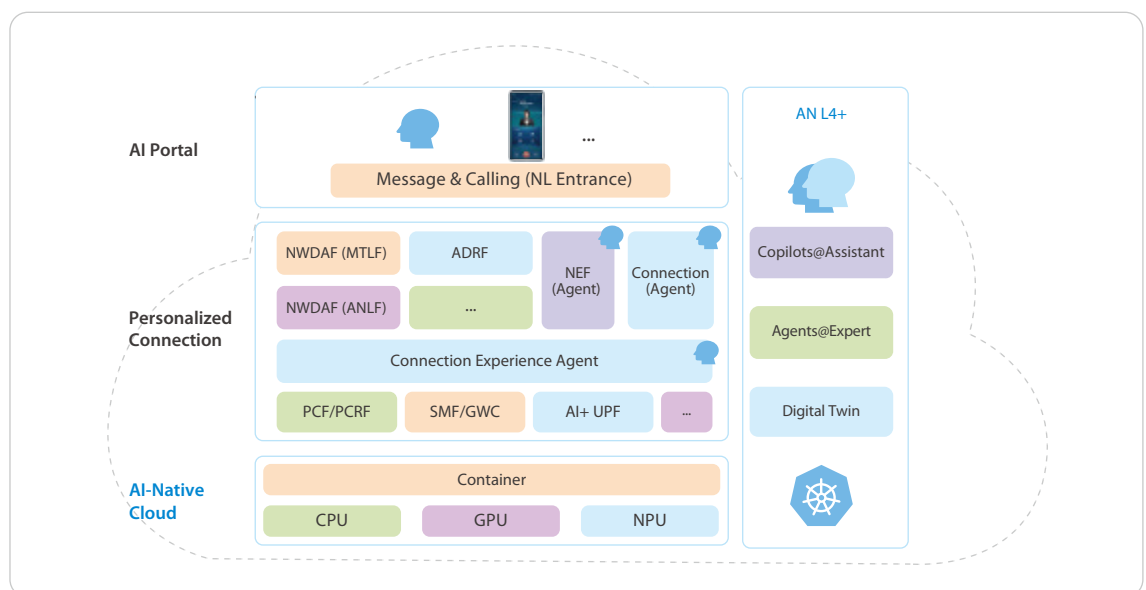
inference frameworks like SGLang, although built on a technology stack independent of the CNCF ecosystem, are deeply integrated with CNCF cloud-native technologies. In K8s environments, mature and production-ready deployment and monitoring practices have been established. Their design naturally aligns with cloud-native principles, enabling seamless integration into existing K8s platforms without modifications.

Thus, K8s has become the de facto standard for managing underlying AI accelerators and serving upper-layer AI models and AI applications. Moreover, as demonstrated by projects such as K8sGPT, K8s is already leveraging AI to enhance cloud platform operations and help O&M teams improve efficiency. Now, the fusion of K8s and AI has positioned K8s as the only available AI-native cloud platform for the ICT industry.

Cloud-Native Telco Cloud: An Inevitable Choice for Intelligent Evolution

To meet the deployment requirements of intelligent network applications and support the intelligent evolution of cloud platforms themselves, building a cloud-native telco cloud platform based on K8s, as shown in Fig. 2, has become a prerequisite and an inevitable choice for intelligent evolution. Especially, the K8s platform built on bare-metal

Fig. 2 Agentic core network architecture.



It is recommended that operators leverage opportunities such as network expansion or new deployments to proactively deploy bare-metal K8s platforms, laying the foundation for intelligent network evolution.

container architectures offers significant advantages.

Smooth Evolution

A K8s platform built with the bare-metal container mode can directly add GPU/XPU servers within existing server clusters to meet the resource needs of intelligent network applications. This not only avoids the need to rebuild the existing cloud platform layer but also enables simple and fast software-hardware integration. Moreover, K8s's native multi-cluster provisioning capabilities can more flexibly meet the complex isolation requirements of network intelligence across different domains of telecom networks. As the most popular cloud-native platform among enterprises, K8s can also better meet operators' future demands to provide 6G-oriented intelligent computing services beyond connectivity.

Convergence and High Efficiency

A cloud-native platform built on K8s inherently provides the convergence of general computing and intelligent computing, along with hybrid scheduling. K8s scheduling can automatically identify AI workloads and allocate the required GPU/XPU resources to them, while allocating general computing resources to other workloads. Meanwhile, KubeVirt, which has entered the production-ready phase, allows K8s to manage virtual machines (VMs) the same way it manages Pods, truly enabling a single platform to manage all resources and provide all services.

End-to-End Autonomy Assurance

Achieving higher levels of network autonomy requires a cloud platform with higher-level autonomy. For example, the network's self-healing capabilities depend

on the cloud platform's capabilities. As an AI-native platform, the integration of K8s with AI further enhances the cloud layer's OAM and intent-based service capabilities, forming a solid foundation for end-to-end higher-level autonomy in telecom cloud networks.

Openness and Decoupling

As the de facto standard for cloud-native platforms, K8s has enabled technology unification across all cloud-native platforms ranging from public cloud to private cloud. Almost all cloud-native platform products and solutions are based on K8s, which is also supported by the most open GPU/XPU hardware ecosystem. This not only gives operators greater flexibility in hardware selection but also guarantees the cross-platform portability of intelligent network applications, promoting openness and decoupling in telecom cloud and avoiding the risk of vendor lock-in.

Currently, the evolution towards intelligent networks has become an industry consensus and a mainstream direction. Most intelligent scenarios and use cases depend on AI empowerment. As the de facto standard for cloud-native platforms, K8s has become the preferred platform for AI accelerator vendors and open-source AI software stacks. Establishing a K8s-based cloud-native platform has become an inevitable path for intelligent network evolution. It is recommended that operators leverage opportunities such as network expansion or new deployments to proactively deploy bare-metal K8s platforms, thereby laying the foundation for intelligent network evolution. **ZTE TECHNOLOGIES**

From AI-Ready to AI-Native: Building a Leading Intelligent Core Network



Wang Weibin

Chief Scientist of
Product Planning,
ZTE



Lu Guanghui

Chief Architect of CCN
Products, ZTE



Currently, AI technology is advancing rapidly, with multiple industries reaching their respective "iPhone moments." From the breakthrough emergence of AIGC to the swift adoption of AI agents and the widespread deployment of AI terminals, AI has moved from laboratories to the industrial applications, igniting a new wave of digitalization and intelligentization.

Against this backdrop, traditional telecommunications networks are undergoing profound transformation and must evolve from being AI-ready to AI-native, shifting from connectivity-centric services to token-centric services, and becoming intelligent hubs for efficient distribution, intelligent routing, and AI services. This transition is not optional but inevitable for competitiveness and growth. As AI advances from a tool to a foundational layer of the network, it will reshape operators' network architectures and business models, creating new opportunities for value creation.

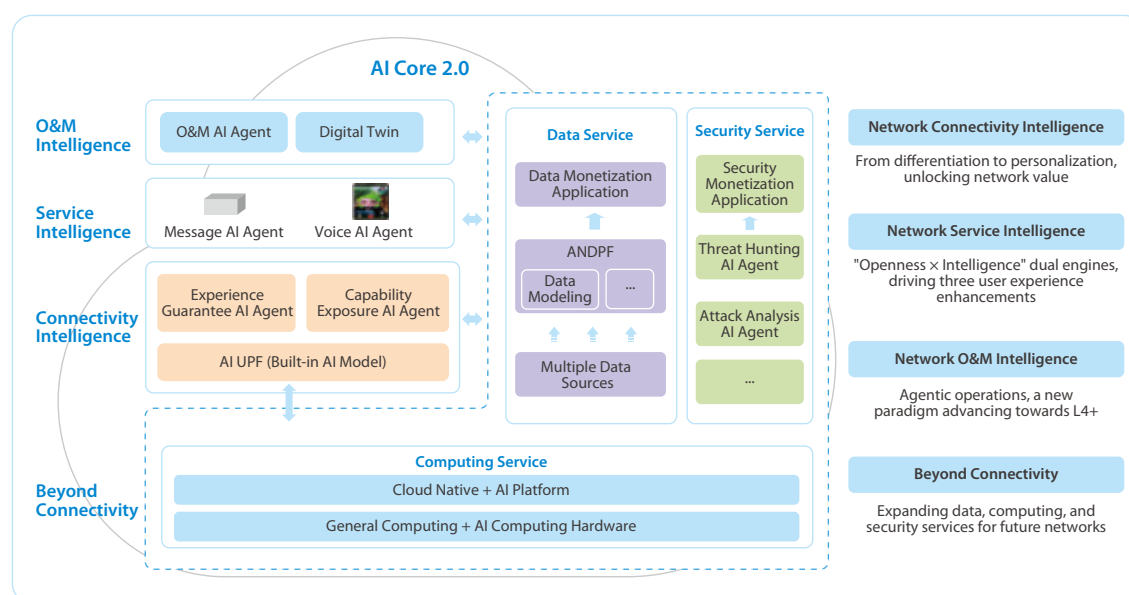
As the hub of mobile networks, the core network bears the critical responsibilities in perception, decision-making, optimization, and operations. Transitioning to an AI-native architecture is key for operators to rebuild their competitive advantage in the AI era. Based on the existing AI Core 1.0 solution, ZTE has launched the AI Core 2.0. Centered on intelligent

agents, this solution systematically redefines core network capabilities in perception, decision-making, and execution—from intent understanding to autonomous evolution—transforming the network into a value engine that enables experience-based monetization, self-healing operations, and scalable ecosystems, meeting the demands of future AI-driven communications.

AI Core 2.0 empowers operators to achieve four key upgrades across four dimensions: network service, network connectivity, network operations and maintenance, and beyond connectivity. Network service intelligence reshapes voice and messaging through enhanced interaction, perception, and terminal entry points. Network connectivity intelligence advances from differentiated connectivity to personalized connectivity. Network operations intelligence builds agent-based O&M through improved interaction, optimization, and prediction, advancing toward L4+ autonomous networks. Beyond connectivity unlocks new growth opportunities by expanding into data, security, and computing services (see Fig. 1).

Network Connectivity Intelligence: Advancing Experience Monetization and Traffic Growth

In the AI Core 1.0 solution, the network



◀ Fig. 1 ZTE's AI Core 2.0 Solution.

connectivity intelligence sub-solution leverages the network data analytics function (NWDAF) on the 5G side as the intelligence core, with PCF serving as the policy orchestration hub and UPF with an embedded AI engine as the execution layer, in coordination with the wireless network to enable end-to-end differentiated connectivity and experience monetization.

To further amplify connectivity value, AI Core 2.0's network connectivity intelligence evolves across three dimensions—experience monetization, traffic boost, and network evolution—driving four major transformations: from differentiated connectivity to personalized connectivity; from reactive service delivery to proactive precision service; from "user seeks service" to "service seeks user"; and from traditional traffic monetization to token-based monetization.

At the experience monetization level, we focus on VIP users and deliver first-class exclusive experiences through AI-powered mobile profiling, cross-domain perception, deep device-network interaction, and experience value management.

At the traffic boost level, targeting mass-market users, we leverage precision marketing and personalized network technologies—such as entitlement-based marketing, green browsing services, video acceleration, and dynamic slicing—to elevate network services from reactive responses to proactive guidance, unlocking new growth opportunities.

At the network advancement level, by introducing token-based charging and NEF+MCP-based composable capabilities, we transform the traditional "data pipe" into a "value exchange platform," fostering a new network ecosystem characterized by intent understanding and proactive value creation.

Network Service Intelligence: A Threefold Evolution of User Experience

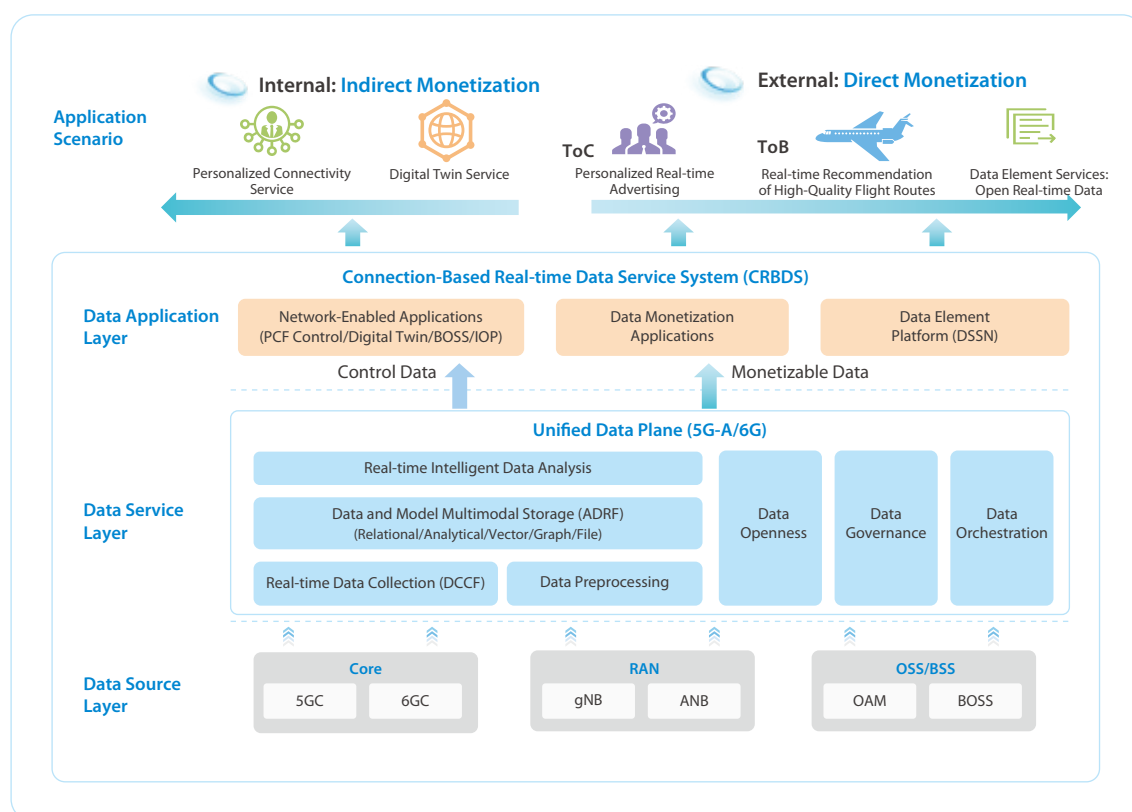
5G New Calling and 5G New Messaging leverage AI to enable network service intelligence, bringing multimodal, interactive, and immersive experiences to real-time voice and messaging services. Network Service Intelligence 1.0 has already achieved large-scale commercial

deployment, while the upcoming 2.0 solution will focus on AI upgrades for voice and messaging. By building an AI Open Platform and deploying two key AI services—vertical AI services for voice and conversational AI agents—this 2.0 evolution will drive threefold upgrades in interaction, perception, and terminal entry points.

- **Interaction upgrade:** Based on the AI Open Platform, plugin-based integration of AR with 5G New Calling enables immersive, intent-driven interaction and spatial intelligence services through the collaboration of AR glasses and AI communication assistants, including a telecom-oriented version of OpenClaw (a conversational agent) and AR-based remote guidance.
- **Perception upgrade:** Through a unified media function (UMF) with built-in AI noise reduction and voiceprint-based noise reduction plugins, call quality in complex environments is enhanced, providing differentiated conversational experiences. Meanwhile, AI-based packet loss compensation technology strengthens media packet loss recovery, delivering a new calling experience in weak network environments.
- **Terminal entry point upgrade:** Built around conversations and driven by AI conversational agent services, a new conversational-agent network architecture is established on the AI Open Platform to facilitate internal and external collaboration, driving the evolution of voice entry points into a smart living portal.

Intelligent Operations: A New AI Agent-Based Operations Paradigm

The traditional "human + tools" operations model cannot meet future requirements for highly efficient operations. How can AI be leveraged to enhance human-computer interaction (HCI), network optimization, and predictive capabilities to achieve efficient operational outcomes? Intelligent Operations 2.0 introduces a new AI agent-based "human + digital workforce" paradigm, supporting the continuous evolution of the core network toward L4+ autonomy. This is primarily reflected in three aspects: easy interaction,



◀ Fig. 2 ZTE's connectivity-based real-time data service (CBRDS) solution.

self-optimization, and predictability.

- **Easy interaction:** The operations model is upgraded from command-driven to intent-driven interaction. Specifically, by introducing multimodal large models, it enables functions such as the generation and verification of intent-based service provisioning scripts as well as knowledge-based Q&A.
- **Self-optimization:** By introducing agent technology, the self-evolution and autonomous capabilities of the operations system are enhanced. This includes comprehensive enhancement of the core "monitoring-diagnose-emergency response" process, enabling automatic fault handling and recovery, and automatic correlation analysis of user complaints.
- **Predictability:** The operations model is upgraded from the after-the-fact passive operations to proactive prediction. By introducing network operation digital twin technology, safer execution of high-risk operations is achieved. Combined with service

digital twin technology, precise prediction and assessment of user experience before service launch are realized, thereby building a highly stable and predictable core network.

Beyond Connectivity: Integrated Computing, Network, AI, Data, and Security Services

Beyond connectivity services, operators unlock new value by leveraging computing networks, with AI as the engine, and delivering data, security, and computing services, transitioning toward token-centric services.

Data Services

AI enhances the value and capabilities of data. According to McKinsey, the potential economic value of data openness (authorized operation of public data) in China reaches RMB 10–15 trillion. Operators can introduce a unified data plane to build a real-time data service system based on connectivity (see Fig. 2), enabling direct data



ZTE will continue to ride the wave of artificial intelligence and mobile network technologies, providing comprehensive support for operators in building AI-native core networks.



monetization through real-time computation, aggregation, AI inference, and open access. It also supports real-time data subscription and aggregation, along with preprocessing and storage of network-domain data, establishing a unified real-time data foundation for network control and data monetization.

Security Services

Built on an integrated cloud-network-security foundation, the solution leverages operators' strengths in ubiquitous connectivity, computing networks, and AI to establish a comprehensive security service system featuring intrinsic security, proactive defense, scenario-specific customization, and compliance assurance. It shifts security from passive protection to proactive immunity and value enablement, addressing three major scenarios: individual users, government and enterprise clients, and industry-specific applications.

Computing Services

The solution builds an AI-native infrastructure integrating general-purpose computing, intelligent computing, and high-performance computing, and enables converged computing and network services. On the one hand, it delivers diversified computing resources to empower scenario-specific AI capabilities in core networks. On the other hand, it introduces AI services at the network edge to provide integrated "connectivity + computing power + intelligence" services.

All-Domain Intelligence Powering Industry Transformation

ZTE has collaborated with leading global operators to launch the AI Core 2.0, an all-domain intelligent solution. It directly addresses the current revenue pressures on telecommunications networks and the challenges of AI deployment, while also laying a solid foundation for the intelligent evolution toward 6G. In network service intelligence, ZTE has released a conversational AI agent to empower voice services, unlocking new value in the communications market. In network connectivity intelligence, ZTE, together with operators, has launched the industry's first intent-driven network operations agent for closed-loop assurance of influencer live streaming. In intelligent operations, ZTE has achieved collaboration between multi-agent systems and network digital twins for the first time, helping operators reduce costs and improve operational efficiency.

ZTE will continue to ride the wave of artificial intelligence and mobile network technologies, providing comprehensive support for operators in building AI-native core networks. Upholding openness, cooperation, and win-win development, it will join hands with industry partners to build a future-oriented intelligent hub that makes connections smarter and networks more intelligent. **ZTE TECHNOLOGIES**

AI-Native 6G Core Network: Shaping a New Paradigm of Intelligent Services

With the rapid development of AI and intelligent agent technologies, network service entities are expanding from humans to intelligent agents with autonomous decision-making capabilities. In the future, 6G networks will evolve from information pipelines for humans into intelligent service platforms supporting efficient collaboration among trillions of intelligent agents. Network evolution will involve three transformations: service traffic transformation, service-mode upgrade, and network-role reshaping.

- **Service traffic transformation:** AI and agent communications will become the dominant source of network traffic, driving demand for new capabilities such as fine-grained service assurance, secure agent identity management, flexible network architecture, and intent-based interaction.
- **Service-mode upgrade:** User requirements are shifting from "communication" to integrated services combining "communication + sensing + intelligence + computing + data + security." The network must become a "beyond-connectivity" service provider with capabilities such as task understanding, intelligent resource scheduling, and collaborative service assurance.
- **Network-role reshaping:** The maturity of large models and agent technologies enables the network itself to evolve into a digital intelligent agent with capabilities such as intent understanding, continuous learning, environmental perception, and autonomous decision-making.

Therefore, the 6G AI core network must build dual capabilities: a "network superhighway" and an "intelligent hub." It serves as both a deterministic infrastructure for massive agent interconnection and a service enablement platform that understands complex intents and achieves intelligent allocation of global resources.

To meet the network demands of the intelligent era, the 6G core network must enable bidirectional empowerment between AI and the network, forming a closed-loop ecosystem of deep AI-network integration and reshaping the value of network services.

- **AI for Network (AI4NET), integrating AI capabilities into the network's DNA:** 6G AI must build upon the 5G network data analytics function (NWDAF), evolving from "add-on" to "native" and from "centralized" to "distributed" to enable atomic-level integration of AI capabilities and network functions and provide the network with native capabilities for self-perception, self-decision-making, self-optimization, and self-evolution.
- **Network for AI (NET4AI), enabling the network to support AI services:** The 6G core network must not only enhance support for AI and agent-based services but also evolve from a connectivity provider into a service provider, offering AI model training and inference services for AI applications, agent interconnection and security management services, and terminal-edge-network-cloud coordinated computing services for business applications.

ZTE proposes an AI-native 6G core network



Zhou Jianfeng

Chief Engineer of
CCN Preresearch
Planning, ZTE



Zheng Xingming

CCN Preresearch
Planning Architect,
ZTE

architecture based on "Intent, Collaboration, Flexibility, and Trust," building a next-generation network infrastructure where communication, sensing, intelligence, computing, data, and security are all delivered as services, driving the evolution of 6G networks beyond connectivity toward a new era of Everything as a Service (XaaS) (see Fig. 1).

Distributed AI Collaboration: From NWDAF to Network-Wide Intelligence

The AI-native 6G core network adopts a distributed AI collaboration architecture to overcome the limitations of the centralized NWDAF architecture in 5G, particularly in terms of accuracy and real-time performance.

- **Distributed embedded AI:** Network functions (NFs) can embed lightweight AI training and inference capabilities on demand, enabling local training and rapid decision-making.
- **Large-small model collaboration:** Lightweight AI models embedded in NFs collaborate with centrally deployed large models to create a hierarchical intelligent architecture that combines agile response with advanced intelligence.

- **Cross-domain coordination:** Core-network AI collaborates with access networks, edge nodes, terminals, and third-party applications to break down data silos and establish unified AI capabilities for network-wide perception, scheduling, and coordination.
- **Multi-agent collaboration:** Multiple agents collaborate autonomously to complete tasks within the core network and across the terminal-edge-network-cloud ecosystem, fostering a collective intelligence environment.

Intent-Driven: From Command-Based Directives to Intent-Based Tasks

Large model technologies empower the 6G network with natural language understanding, enabling users to customize service experiences through dialogue and lowering interaction barriers. The AI-native 6G core network will provide three intent entry points:

- **User-oriented intent entry:** End users can convey their needs in natural language. The network automatically interprets the intent and completes end-to-end configuration, including QoS

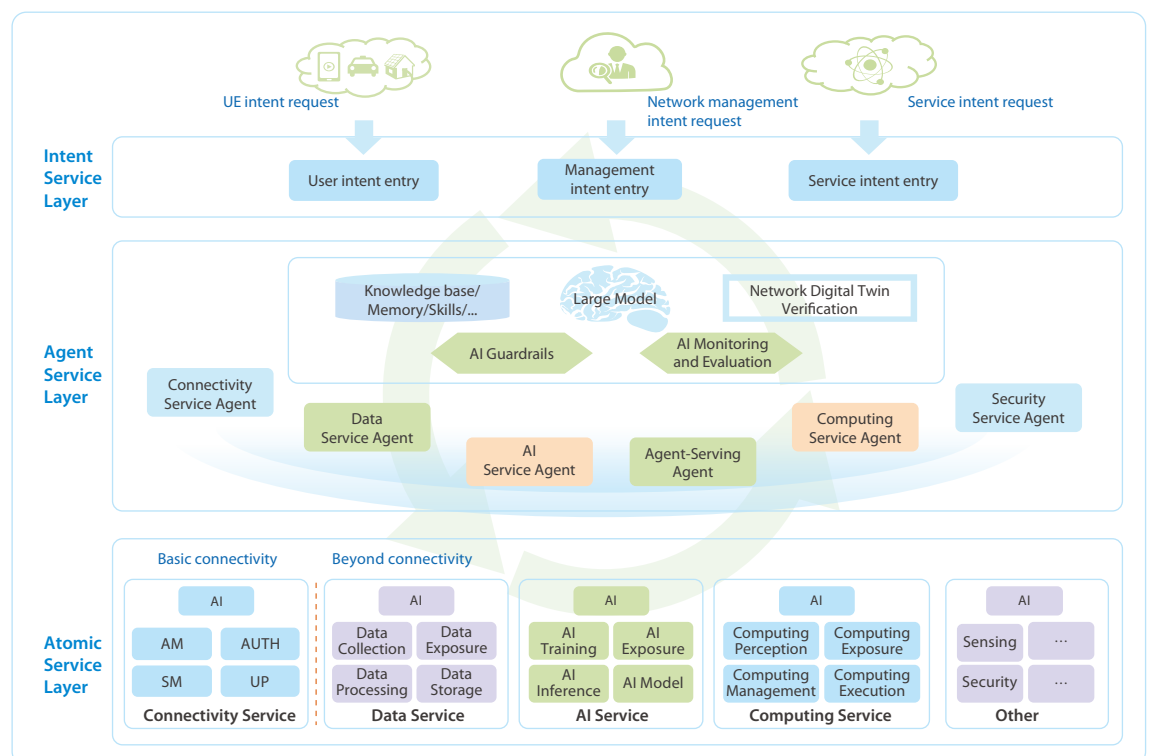


Fig. 1 AI-native 6G core network architecture.

configuration and computing resource allocation.

- **Management-oriented intent entry:** O&M personnel can achieve zero-touch network management through intents, establishing a closed-loop autonomous system covering intent, planning, configuration, and validation, thereby reducing operational complexity.
- **Service-oriented intent entry:** Third-party applications can access network capabilities and information through intents, enhancing innovation efficiency and simplifying service development.

Multi-Dimensional Service Innovation: From "Connectivity Pipelines" to "Service Hubs"

The AI-driven 6G core network will evolve beyond the traditional "pipeline" model into an integrated service platform offering Network as a Service (NaaS), Sensing as a Service, AI as a Service (AlaaS), Computing as a Service (CaaS), Data as a Service (DaaS), and Security as a Service (SECaaS).

- **Network services:** Building upon foundational capabilities such as communication forwarding, QoS guarantee, and inter-networking, the network supports on-demand service customization to deliver personalized experiences.
- **Sensing services:** The network provides services such as object detection, high-precision positioning, and integrated sensing and communication, empowering physical-world intelligence.
- **AI services:** The network offers AI training and inference services, AI model management services, and agent services, enabling intelligent service innovation.
- **Computing services:** Leveraging multi-dimensional awareness across communication, sensing, intelligence, computing, and security, the network enables intelligent orchestration and globally optimized scheduling of computing resources.
- **Data services:** The network provides value-added services such as real-time data analysis, model hosting, and personalized profiling, facilitating data value monetization.
- **Security services:** The network provides elastic

security capabilities such as identity authentication, permission control, risk detection, and privacy protection for networks, data, AI, and applications.

Flexible Programmability of Capabilities: From Static Configuration to Dynamic Service Orchestration

With an agent-based architecture, network functions are no longer fixed but are orchestrated and dynamically combined on demand.

- **Functional atomization:** Network functions and services are decomposed into orchestratable microservice units, which are combined on demand by AI agents to form customized network logic.
- **Network service programmability:** Network functions evolve into agent-based entities with flexible reconfiguration capabilities, enabling rapid adaptation to personalized scenario requirements and enhancing service innovation efficiency and user experience.

Agent Services: From Transparent Forwarding to Service Proxying

In response to the explosive growth of agent-based applications, the 6G core network needs to provide enhanced service capabilities for agents.

- **Agent identity management:** Assigns globally unique identifiers to AI agents, supporting identity authentication and trust evaluation.
- **Agent registration and discovery:** Supports AI agent registration, publication, and discovery, enabling inter-agent connectivity and collaboration.
- **Dynamic agent networking:** Dynamically creates AI agent subnets and enables secure interworking and adaptation among AI agents across vendors and platforms.
- **Personal agent proxy:** Serves as the hub for AI agent services, extending beyond basic functions such as forwarding and assurance to act as a proxy for user intent-based tasks, including task decomposition, execution, and validation.

Through bidirectional empowerment via AI4NET and NET4AI, the AI-native core network is evolving into an "intelligent hub" capable of intent understanding, capability orchestration, service innovation, and trustworthy AI.

AI Self-Evolution: From Fixed Capabilities to Continuous Growth

Through continuous learning and self-evolution, the 6G core network can learn the operational patterns of the local network, automatically update its domain knowledge base and skill sets, and better adapt to local network characteristics and service requirements.

- **Online learning:** The network continuously collects data during operation and optimizes model performance through incremental learning, reinforcement learning, and other methods to adapt to dynamic network and service changes.
- **Policy self-optimization:** Network agents can evaluate policy effectiveness, autonomously propose optimization recommendations, and implement them after validation in a digital twin environment.
- **Knowledge accumulation and inheritance:** The network builds long-term memory systems, such as knowledge bases and skill repositories, to enable knowledge accumulation, cross-domain sharing, and long-term self-evolution.

AI Trustworthiness, Reliability, and Evaluability: From Functional to User-Friendly

AI trustworthiness and reliability are critical for mobile networks. The 6G core network must establish a trust assurance system across the full AI lifecycle to ensure that AI behaviors are transparent, controllable, and accountable.

- **Ingress security screening:** Performs semantic

analysis and risk identification on user intent to block malicious directives and unauthorized requests.

- **Egress security fence:** Validates AI inference results for compliance and enforces boundary constraints to prevent unauthorized decisions and abnormal outputs.
- **End-to-end process monitoring:** Provides real-time monitoring of the entire "Intent–Decision–Execution–Feedback" process, ensuring process visibility, auditability, and traceability.
- **Trustworthy evaluation of inference results:** Establishes an AI performance evaluation system to quantify the accuracy of AI inference results and enable risk pre-alarm and other assessments.
- **AI abnormal rollback:** When an AI module fails, the system can bypass the AI functionality and revert to local algorithms or baseline functions to ensure stable system operation.

The 6G core network is undergoing a paradigm shift from "connectivity" to "service." Through bidirectional empowerment via AI4NET and NET4AI, the AI-native core network is evolving into an "intelligent hub" capable of intent understanding, capability orchestration, service innovation, and trustworthy AI. It can deliver high-value external services, such as AI agent communication, personal AI agent proxy, and computing resource scheduling, enabling the 6G digital society. ZTE will continue to advance the native integration of AI and networks, collaborating with industry partners to build a new paradigm of intelligent services. [ZTE TECHNOLOGIES](#)

Redefining Experience Operations to Unlock Network Monetization

Communication networks have transformed from voice-centric services to data traffic, and now to AI-driven stage. In the traffic-plus-AI operations phase, a primary challenge for operators is how to increase user ARPU. To address this, experience operations are evolving from version 1.0 to 2.0. In the 1.0 era, the industry focused on basic network coverage and traffic capacity provisioning, achieving preliminary exploration of experience operations. In the 2.0 era, the core logic has shifted towards precision, intelligence, and a scenario-based approach. Through deeper cross-domain collaboration on the wireless side, iterative evolution of user profiling large models, and enhanced capabilities of experience assurance agents, operators can break through 5G monetization bottlenecks and build a new closed-loop business model driven by user experience (see Fig. 1).

Precision-Driven Collaborative Enhancement Reshapes Network Experience Assurance

In the Experience Operations 1.0 phase, wireless-side assurance was largely confined to a single dimension, and multiple pain points hindered the realization of experience value. Today, cross-domain collaboration has achieved two major breakthroughs, laying a solid network foundation for Experience Operations 2.0.

For dedicated bearer assurance, the drawbacks of the previous best-effort assurance model have become prominent—once a base station cannot meet demands due to limited resources, insufficient coverage, or interference, the subscribed user experience is directly affected.

However, through intelligent perception enabled by wireless collaboration, an end-to-end cross-domain assurance system can be established. Relying on the evaluation capabilities of the wireless side, the system can provide feasibility assessments for dedicated bearer assurance. For example, when a user experiences stuttering during gaming, wireless evaluation conclusions can help determine whether an additional assurance package is needed, precisely resolving network issues. The wireless-side default bearer rate-tier evaluation leverages AI to assess different user classes at the base-station level and accurately report the maximum rate available for each tier at the user's location. Users can view their real-time uplink and downlink rates by tier and independently determine whether to upgrade their plans for an improved experience.

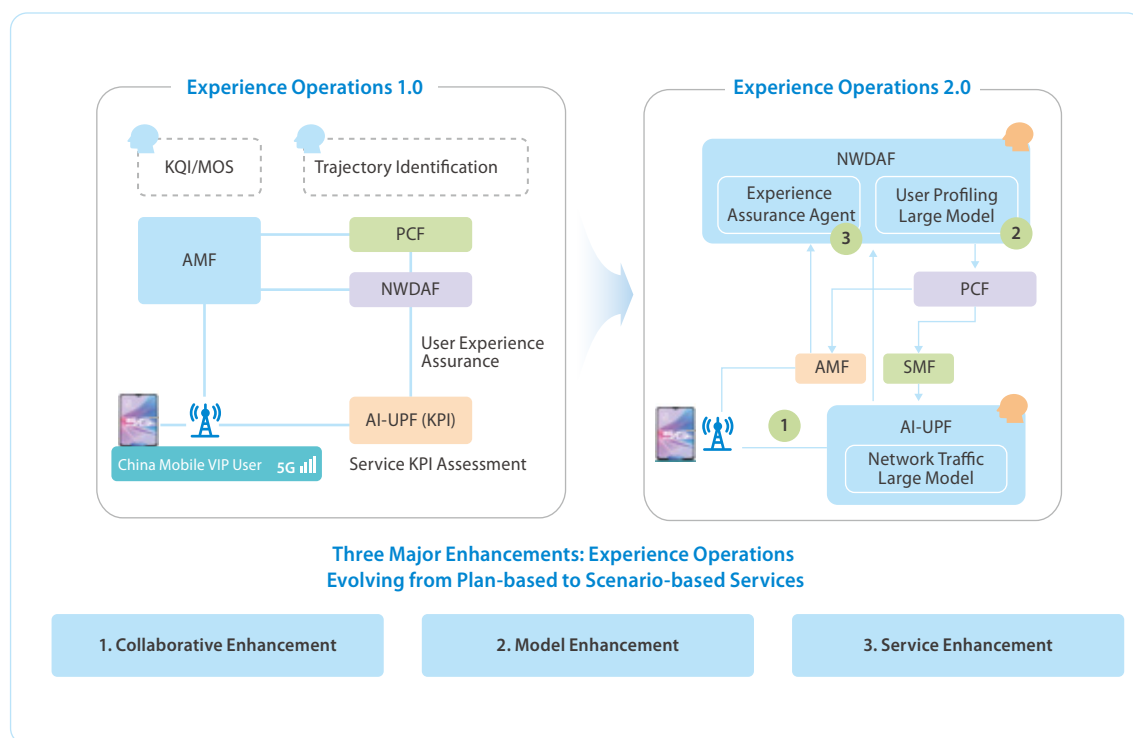
Traditional slice replacement solutions rely on terminal support for UE route selection policy (URSP), making them difficult to scale and unsuitable for meeting the needs of large-scale scenarios. Through deep collaboration with the wireless side, the new solution triggers user-level slice replacements based on time, location, and application experience. Although service flow-level replacement is not yet supported, the solution eliminates dependence on terminals, significantly lowers deployment barriers, and can be quickly applied to diverse scenarios such as smart offices and HD live streaming. It provides enterprise customers and individual users with stable, customized network slice services. This cross-domain collaboration capability enables the network to shift from "passive adaptation" to "active response," laying the foundation for the design and monetization of differentiated experience products.



Cao Yilin

Chief Engineer of CCM
Product Planning, ZTE

Fig. 1 Evolution of experience operations toward scenario-based services.



Intelligent Model Enhancement Powers Scenario-Based Monetization

User profiles are a core lever for experience management, and their accuracy directly determines monetization efficiency. In the era of Experience Management 1.0, user profiling was largely based on time-series models of subscriber-group mobile signaling, featuring limited dimensions and weak scenario adaptability. In the 2.0 era, these models have evolved toward spatio-temporal modeling—integrating information such as subscriber group signaling and app behaviors, and incorporating user mobility sequences and the spatial associations between users and base stations—marking a leap from "shallow recognition" to "deep insight."

After the upgrade, the large model gains strong scenario generalization capabilities, enabling accurate identification across diverse settings such as stadiums, campuses, tourist attractions, and subways. It establishes a multi-scenario role recognition system that defines identities for staff, scene users, and other groups based on unique behavioral patterns.

In stadium scenarios, the model precisely identifies target users by analyzing user location and service usage data, enabling actions like plan recommendations, benefit assurance, and logo display to deliver targeted services.

In campus scenarios, it leverages students' regular routines and academic cycle patterns to accurately identify student groups, matching them with exclusive data plans and campus-specific green services.

In tourist attraction scenarios, the model captures seasonal variations in visitor routes, clearly mapping sightseeing routes and staff patrol trajectories. By analyzing the mobility patterns of visitors entering the attraction in real time, the system pushes service cards to those moving at a slower pace, encouraging them to open the attraction's app and collect interactive badges, enriching their experience.

In subway scenarios, it detects peak-hour tidal patterns and in-station commercial ecosystems, precisely distinguishing commuters, staff, and station merchants, and assigning differentiated network services and commercial benefits.

This precise, multi-scenario profiling capability

As technology continues to evolve, experience operations will evolve from traffic monetization and explicit experience indicators toward more intelligent, contextual, and precise service delivery.

transforms service delivery from a broad, one-size-fits-all approach to a targeted, precision-driven model, significantly improving user conversion rates and retention.

Scenario-Based Service Enhancement Ushers in a New Era of Interactive Operations

The ultimate goal of experience assurance is to achieve "perceivable experience, convertible value." The evolution of experience assurance agents has made this goal a reality. In Experience Operations 1.0, experience assurance primarily relied on subscription-based services and OTT-enabled services, which featured high adoption barriers and indirect user perception. As a result, its core focus was limited to a broad "network-to-user" model that struggled to meet personalized needs.

In the 2.0 era, experience assurance agents have undergone comprehensive upgrades. With intent-based interaction and recognition at their core, they enable precise workflow orchestration, allowing the resulting workflows to invoke atomic capabilities as planned and continuously optimize execution.

This upgrade delivers two core values: on the one hand, it significantly lowers the barrier to capability adoption, making the deployment and iteration of experience assurance more efficient; on the other hand, it enables more precise user experience perception at the

subscriber end, driving a transformation from passive network-driven service ("network finds user") to proactive, interactive service ("user finds network"). Users no longer need to wait passively for network optimization; instead, they can actively trigger assurance services based on their needs, such as enabling low-latency protection during HD live streaming or one-tap activation of high-speed service in office scenarios. This interactive assurance mode not only enhances customer satisfaction but also expands the boundaries of experience monetization, turning experiential value into tangible business gains.

The evolution from 1.0 to 2.0 represents a deep alignment between technological innovation and user requirements. Cross-domain collaboration on the wireless side resolves key pain points in network assurance; user profiling large models precisely capture contextual needs; and intelligent experience assurance redefines service interaction patterns. Together, these three core enhancements unlock the full potential of the network and establish a complete "Experience-Assurance-Monetization" ecosystem.

In the future, as technology continues to evolve, experience operations will evolve from traffic monetization and explicit experience indicators toward more intelligent, contextual, and precise service delivery, advancing further toward token-based operations and creating broader revenue opportunities across consumer markets and vertical industries. **ZTE TECHNOLOGIES**

AI Calling: A Strategic Elevation from Voice Portal to AI Life Portal



Lu Jiwen

Chief Engineer of CCN
VoiceMSG Product
Planning, ZTE

As user scale and average call duration plateau, ARPU growth in operators' voice services is reaching a bottleneck. Intensifying competition and continuous erosion from OTT services are also pushing the voice business toward a structural inflection point. While traditional voice networks have the natural advantage of being a natural voice portal, operators must move beyond duration-based management, redefine the value of voice networks, and transform voice from a passive tool into an active service, shifting from a "voice portal" to an "AI life portal" and enabling a new value-based management model.

To support this shift, ZTE has launched an AI Calling solution powered by network-native AI (see Fig. 1). Built on three capability pillars—domain-specific intelligence, an open platform, and conversational agents—the solution reshapes the voice service value chain, drives ARPU growth, and redefines ecosystem value.

Voice Domain-Specific Intelligence

Addressing challenges such as weak signals in complex scenarios, high background noise in public places, scam calls, hearing barriers, and language limitations, ZTE has developed voice domain-specific intelligent services. These include AI-powered ClearTalk for enhanced call clarity, AI anti-fraud for millisecond-level risk detection and real-time alerts with over 95% accuracy, intelligent translation for real-time multilingual speech-to-speech communication, and sign language calling that combines video streaming and AI gesture recognition to provide a full-duplex interaction channel for hearing-impaired users.

ZTE pioneered ClearTalk, enabling large-scale

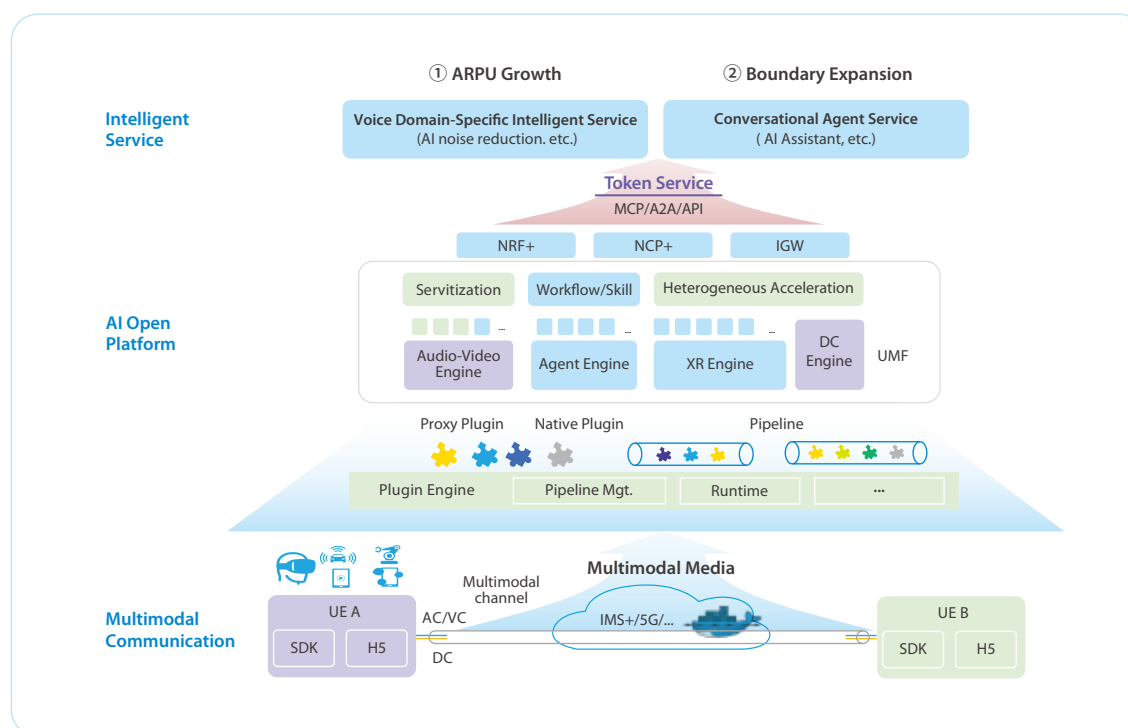
deployment of network-native AI noise reduction and AI weak-network compensation technologies. By integrating voiceprint recognition, ambient noise suppression models, and AI-PLC models into the core network media plane, the solution enables coverage across all terminals, scenarios, and networks, delivering real-time and precise voice enhancement, filtering out ambient noise and nearby voices, and improving packet-loss compensation capability by more than five times. Field tests show that in high-noise, weak-coverage environments, speech clarity increased to over 98%, call disconnect rates decreased by 42%, and speech recognition accuracy improved by 35%. This technology requires no terminal upgrades and can be rapidly deployed via plug-in modules.

AI Open Platform

The AI Open Platform complies with 3GPP standards and is built on the IMS network. Through seamless upgrades, it transforms traditional voice networks—once closed, rigid, and limited in service scope—into intelligent, open, orchestratable, and rapidly evolving "AI+" networks.

The platform consists of four components: the unified media plane (UMF), the new calling platform (NCP+), the interworking gateway (IGW), and the enhanced network repository function (NRF+). Together, they enable dynamic service orchestration, on-demand tool loading, and capability exposure, providing full-stack AI enablement for voice networks.

- UMF supports "plug-and-play" AI functions, allowing dynamic loading and orchestration of noise reduction, fraud detection, translation, sign language recognition, and third-party intelligent agents.



◀ Fig. 1 Architecture of the AI Calling solution.

- NCP+ provides service-level APIs for open capability exposure, enabling the dynamic creation of new composite voice services, such as activating communication assistants during calls to initiate multi-party real-time conversations, plan itineraries, schedule meetings, and generate call summaries.
- IGW collaborates with intelligent agents across industries via MCP, A2A, and other protocols to enable cross-domain integration of CT and IT services. For example, the conversational agent can serve as a plug-in tool within OpenClaw.
- NRF+ supports dynamic registration and discovery of intelligent agents and tool plug-ins.

Conversational Agent

AI Calling integrates communications, AI, and services to reimagine human communication and realize Communications-as-a-Service. ZTE's conversational agent delivers carrier-class OpenClaw capabilities, providing end-to-end intelligent services before, during, and after calls. During a call, users can invoke tool plug-ins from the Open Platform with one click to enable service provisioning, multimodal

interaction, content delivery, and application sharing. Embedded within the call, the conversational agent supports real-time semantic understanding, intent recognition, and task execution, enabling an intelligent closed loop of "speak once, accomplish once."

The closed-loop "Communication-AI-Service" design transforms traditional voice networks from a simple calling tool into an intelligent service platform that engages users, unlocks new scenarios, and creates value, driving breakthroughs in both service usage and business models.

AI Calling is a paradigm shift in communication services, enabling operators to evolve from a "voice portal" to an "AI life portal." Built on a network-native AI kernel, an open platform foundation, and multimodal intelligent interfaces, it establishes a new value system. During the transition from 5G to 6G, operators leading in intelligent conversations will not only retain their voice entry point but also shape the future gateway to intelligent services. ZTE will empower operators to transition from "duration-based management" to "value-based management," ushering in a new era of intelligent communication. **ZTE TECHNOLOGIES**

“Car-Home Connect”: A Pan-Domain Private Network for People, Vehicles, and Homes



Liu Xiliang

Chief Engineer of
5GC Product Planning,
ZTE



Si Hairui

Technical Manager of
5GC Product Planning,
ZTE

As digital infrastructure enters the AI-native era, 5G-Advanced and general artificial intelligence are converging to move connectivity beyond data transmission toward cognitive network services. Ubiquitous intelligent connectivity is becoming a digital neural hub that integrates perception, connectivity, computing, and intelligence, enabling seamless collaboration among humans, machines, and things, as well as real-time interaction between digital twins and the physical world.

However, fragmented infrastructure, heterogeneous computing networks, and blurred data sovereignty boundaries are hindering further digital development. Building an intelligent connectivity foundation based on a unified protocol stack and a trusted cross-domain computing architecture has therefore become a strategic priority. “Car-Home Connect” is an intelligent network service solution designed for this era, overcoming ecosystem lock-in and enabling cross-industry collaboration.

Car-Home Connect: Elevating Pan-Domain Private Network Value

Car-home Connect enables the evolution from functional connectivity to private network services by establishing a unified technological foundation that connects ToC, ToB, and ToH multi-domain services. Its core lies in redefining connectivity: from simple data flows to the lossless delivery of intent; from platform dominance to the restoration of user sovereignty.

Built on open network services and aligned with industry standards, the Car-Home Connect personal pan-domain private network supports cross-domain protocol adaptation, enabling “barrier-free dialogue” among various devices. Ultimately, it forms a human-centered, trusted, and co-creative intelligent digital ecosystem: homes become emotional living spaces, vehicles become mobile intelligence nodes, and industry services become on-demand external intelligence. This enables an integrated system of people, vehicles, and homes that supports the borderless flow of value within a trusted private network (see Fig. 1).

● Consumer-Side Connectivity: Seamless Interconnection Beyond Ecosystem Lock-in

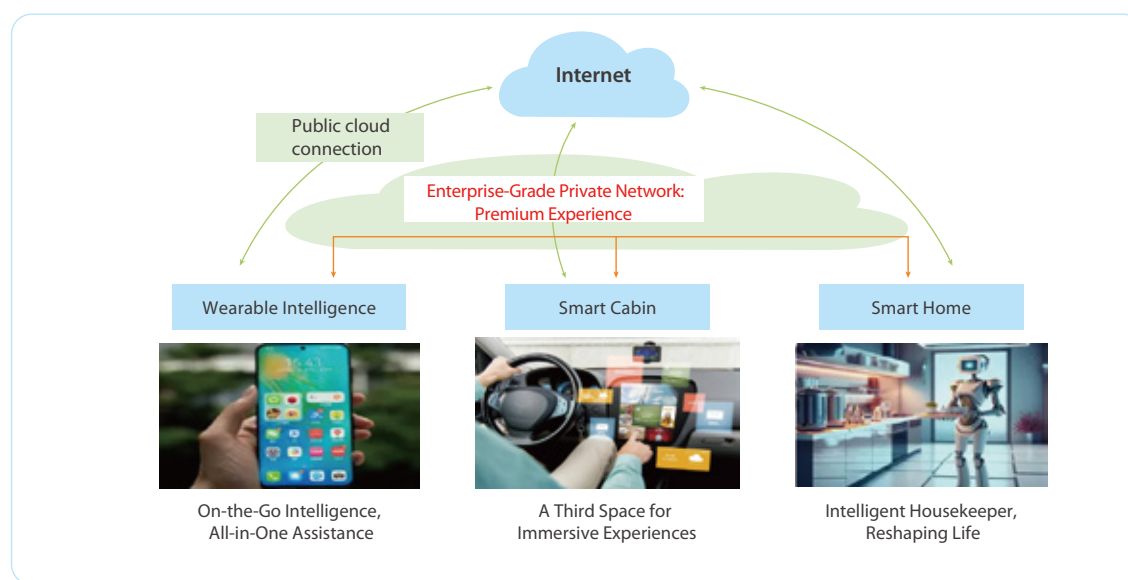
Seamless cross-device access among smartphones, in-vehicle systems, and home automation devices; seamless flow of data and services across domains; unified identity management within private networks, low-latency local forwarding, and elimination of vendor lock-in.

● Industry Connectivity: Opening New Service Entry Points and Reshaping the Value Chain

Serves as a critical channel for vertical industries to reach home users; provides open cross-domain interfaces for vendor-customized services and helps break platform dominance over traffic.

● Home Intelligence Hub: Secure, Controllable, and User-Configurable

Multi-layer isolation technologies ensure home data security; and users can autonomously configure and organize their networks, redefining the paradigm of home network management.



◀ Fig. 1 Pan-domain private network enables reliable and secure network services.

Key Technologies of the Car-Home Connect Pan-Domain Private Network

The Car-Home Connect pan-domain private network solution marks the evolution of communication networks from connectivity to services. Driven by demands across experience, technology, industry, and security, it provides the core infrastructure for future personal and household digital life. The solution breaks down data silos, bridges gaps in cross-scenario experiences, and introduces 6G-oriented technologies into 5G networks to accelerate the deployment of intelligent-agent networks.

Centered on mobile networks, Car-Home Connect integrates intelligent agent technologies, cross-domain interworking, and unified identity authentication and management to enable seamless connectivity across ToC, ToB, and ToH domains. The architecture is illustrated in Fig. 2.

Intelligent Agent Technology

In the Car-Home Connect pan-domain private network, agents serve as an "invisible hub" connecting people, vehicles, and homes, seamlessly integrating three isolated scenarios and redefining personalized network services.

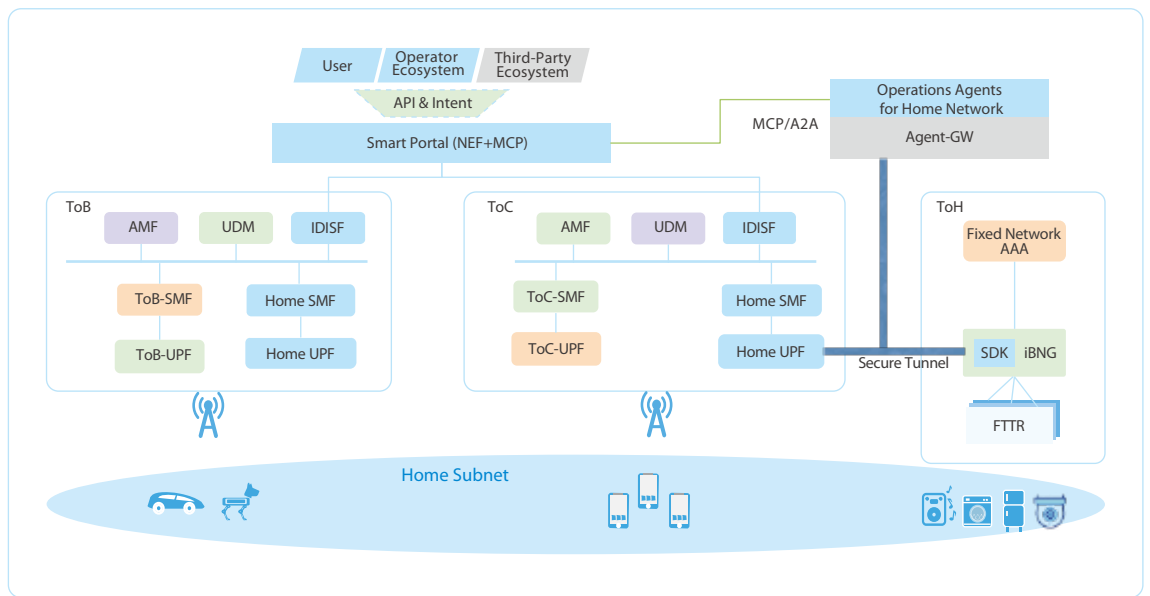
Endpoint agents, designed for end users, act as the entry point for network services and

applications, enabling cross-domain coordination through intent-driven interaction—users need only issue a single directive to control smartphones, in-vehicle systems, and smart appliances, without separate operations. For example, upon hearing the command "Heading Home," the endpoint agent interprets the intent, coordinates with network-side agents to set up the communication network and connect relevant smart devices. It also activates car functions such as navigation, seat adjustment, and driving assistance, while proactively turning on the home air conditioner and adjusting the lighting.

The network-side collaborative agent primarily consists of the capability exposure agent and the inter-domain integrated service function (IDISF) agent. The capability exposure agent serves as the central hub for capability exposure across the private network, breaking down barriers between devices, networks, and ecosystems across people, vehicles, and homes through AI-driven autonomous decision-making and standardized capability encapsulation. It encapsulates dispersed capabilities—such as vehicle control, home automation, 5G/6G network capabilities, and AI perception—into schedulable and accessible standardized services, serving as a key enabler of seamless access to the pan-domain private network and network-service integration.

The inter-domain integrated service agent focuses

Fig. 2 A pan-domain private network built around agents.



on cross-domain coordination and holistic network control. It is responsible for inter-domain connectivity, unified allocation and management of device addresses across domains, trusted authentication and access control for users and devices, as well as secure resource scheduling and traffic orchestration. It provides a stable, secure, and efficient network infrastructure that supports cross-domain collaboration of the capability exposure agent and seamless connectivity across people-vehicle-home scenarios, acting as the core hub of the pan-domain private network solution.

Inter-Domain Interworking Technology

Inter-domain interworking technology is built on overlay network technology to establish a unified cross-domain IP management system, enabling flexible IP-layer interconnectivity for people-vehicle-home services. Home devices are unaware of the overlay-layer addresses; only other network elements within the subnet—such as the portal, IDISF, integrated broadband network gateway (iBNG), home SMF, and home UPF—can perceive and use these addresses for service interactions. This architecture ensures flexible cross-domain interconnectivity for people-vehicle-home services without requiring modifications to home devices, thereby reducing

deployment complexity and O&M costs.

- **IDISF:** Serves as the core node for overlay address allocation. It is responsible for uniformly assigning private subnet IP addresses to mobile and home devices, ensuring the uniqueness and manageability of these addresses.
- **Core SMF&UPF:** Reuse the existing NAT solution for dual-domain private network IP addresses to enable translation between public and private network addresses.
- **Home SMF&UPF:** Introduce a new NAT capability to map between home private IP addresses and people-vehicle-home subnet IP addresses, enabling interconnectivity between the home private network and the overlay service layer.
- **iBNG:** Interacts with the mobile core network to proactively obtain overlay-layer IP addresses of people-vehicle-home subnets for various home devices, as well as the corresponding mappings to private IP addresses within home networks, providing the address information required for cross-domain service scheduling.

Unified Identity Authentication Management System

Mobile networks and home broadband devices belong to different security domains and administrative domains. To achieve cross-domain

The construction of the Car-Home Connect pan-domain private network is not only a scenario-driven implementation of 5G technology, but also a profound transformation of digital lifestyles.

interconnection, the core challenge is identity interworking between different systems. Therefore, the solution introduces a unified home subnet identifier, HomeNetID. Through unified subscription at service halls or via operator apps, HomeNetID is associated with mobile phone numbers and home IP address segments, serving as a cross-domain, cross-subnet identifier that links people, vehicles, and homes. Meanwhile, based on unified data management (UDM), the solution uniformly manages home subscription information and extends home subscription data without altering the existing subscription data structure, enabling unified authentication of home gateways in the mobile network through their digital IDs and corresponding security keys.

The solution also supports flexible subscription for households, enabling each user to be associated with up to X home subnets and each home subnet to connect to up to Y home gateways. Through cross-domain identifier association, user equipment identifiers in mobile networks, such as IMSI, MSISDN, and IMEI, are associated with the HomeNetGroupID, HomeNetID, and the corresponding digital IDs and security keys of home gateways. This establishes a clear subscription information structure in the mobile network and provides trusted identity management and security support for cross-domain interconnection.

Future Evolution: From Scenario Interconnection to Ecosystem Prosperity

The construction of the Car-Home Connect pan-domain private network is not only a scenario-driven implementation of 5G technology,

but also a profound transformation of digital lifestyles. Technologically, it drives 5G evolution from "wide coverage" to "deep penetration," releasing the value of 5G technology through refined operation of private networks. Industrially, it breaks down barriers among closed ecosystem vendors, providing small and medium service providers with direct access to users and fostering integrated innovation across "home + vehicle + industry" scenarios. From the user perspective, it brings "full-scenario smart living" from concept to reality, redefining family life in the digital era with private, personalized service experiences.

With the accelerated integration of AI technology and 6G networks, the personal pan-domain network will continue to evolve along two directions:

- **Proactive intelligence:** Deep integration of IDISF with large models enables proactive prediction of user intent and seamless triggering of services.
- **Ubiquitous coverage:** 6G space-air-ground integrated networks further extend the coverage of private networks, allowing users to enjoy a consistent and intelligent service experience in outdoor and high-speed mobility scenarios.

At the convergence of 5G and 6G, the exploration of personal pan-domain networks not only provides application scenarios for current technologies, but also offers valuable insights for the scenario-driven design of future networks, driving seamless interconnection across people, vehicles/machines, and homes, and advancing it from functional availability toward superior user experiences, while fostering a safer, smarter, and more open digital life ecosystem. **ZTE TECHNOLOGIES**

Core Network O&M Agents: Enabling Continuous Evolution of L4 Autonomous Networks



He Wei

Chief Engineer of CCN
Intelligent O&M
Product Planning, ZTE

The evolution from L1 to L5 in autonomous networks represents a fundamental transformation in O&M productivity and operating models. As the industry moves beyond the 'human-led, tool-augmented' phase toward advanced L4 autonomy, the key challenge is achieving truly autonomous closed-loop operations. O&M agents are reshaping autonomous closed-loop operations in core network O&M by enabling a shift from "human-operated, system-assisted" to "human-defined intent, agent-driven closed loops," integrating digital employees into daily O&M activities.

Three Key capabilities Reshaping Core Network O&M

Based on TM Forum's overall architecture evolution roadmap and ZTE's O&M experience, ZTE's intelligent core network system, OpsAgentStudio, deeply integrates key technologies such as agents and digital twins (DT) into the O&M framework. It delivers three critical capabilities: easy interaction, shifting O&M from command-driven to intent-driven; self-optimization, enabling agents to autonomously reason, analyze, locate, and resolve faults in complex scenarios; and predictability, leveraging digital twin technology to simulate changes and forecast risks in a virtual environment, enabling a transition from reactive response to proactive validation.

The architecture of ZTE's intelligent core network O&M system is shown in [Fig. 1](#).

Easy Interaction

In traditional O&M models, human-system

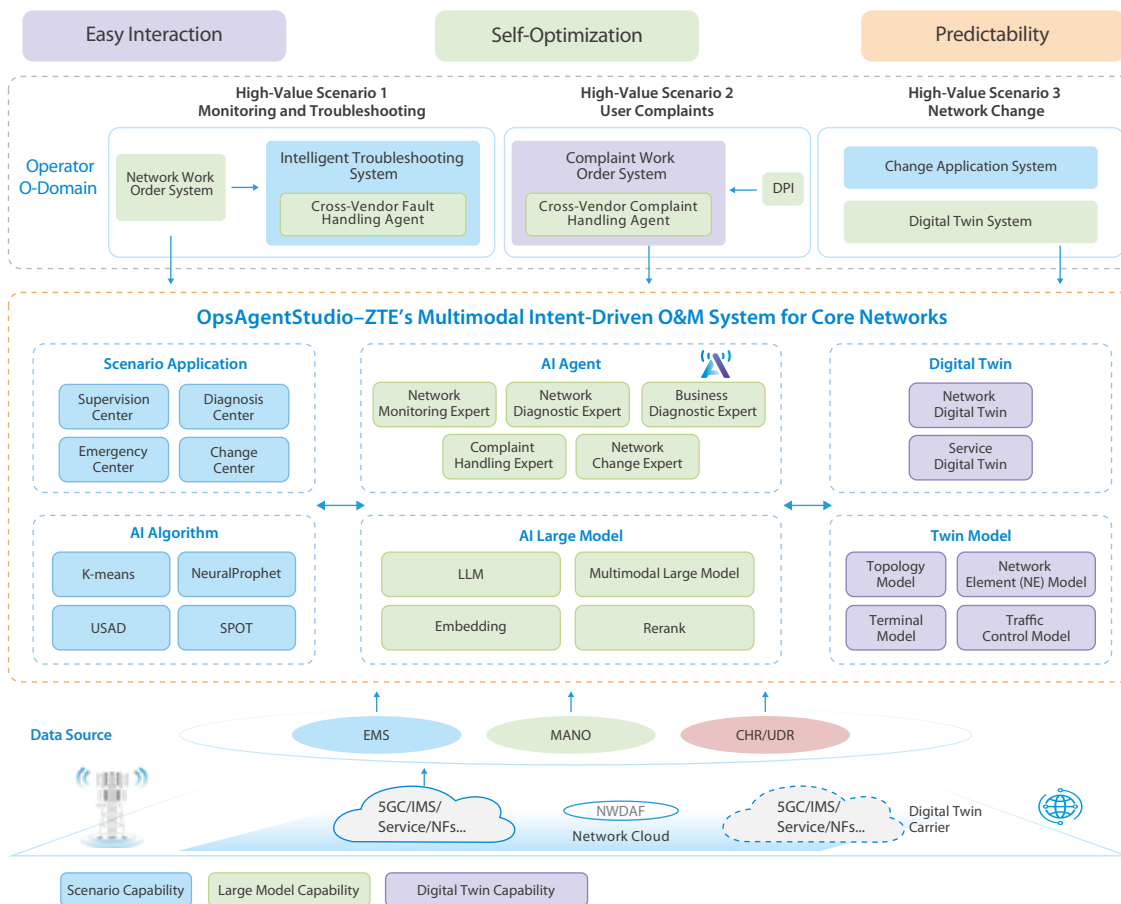
interaction is command-based. Engineers must be proficient in the workflows and command syntax of various network management systems and often switch between tools to execute repetitive workflows. This approach is inefficient and time-consuming.

Intelligent agents transform this model. In intent-driven O&M, operators only need to express the desired outcome, while agents determine how to achieve it. When network abnormalities occur, operators no longer need to troubleshoot alarms or execute commands one by one. Instead, they submit intents to the agents through natural language or standardized interfaces. The agent then completes the closed-loop—from goal decomposition, data collection and root cause analysis to solution generation, operation execution, and result validation—with human confirmation required only for critical decisions or abnormal situations.

"Easy interaction" lowers operational barriers and enhances human-agent collaboration. Humans focus on business decisions, change approval, and complex scenario analysis, while agents handle real-time processing of massive data, repetitive operations, and standardized decisions. As operators shift from "tool users" to "agent managers," interaction costs decrease and response speed improves.

Self-Optimization

Traditional automated systems follow deterministic "condition-action" logic and can only handle predefined scenarios. When facing cross-domain correlated failures, unknown anomalies, or dynamically changing network environments, these systems often fail, still requiring



◀ Fig. 1 Architecture of the intelligent core network O&M system.

human intervention. Intelligent agents overcome this limitation with autonomous reasoning and analytical capabilities, enabling true self-optimization.

The value of self-optimization lies in transforming operational expertise from human memory into system-embedded knowledge. By combining large language models for semantic understanding, knowledge graphs for relational reasoning, and reinforcement learning for decision optimization, agents can better understand the complex core network operational states. Moreover, through continuous feedback from fault handling and network adjustments, they improve over time, becoming digital O&M employees that grow smarter with use.

Predictability

A key hallmark of L4 advanced autonomy is the shift from reactive response to proactive anticipation. In core network operations, configuration changes are among the highest-risk

activities, where a single misconfiguration or improper routing policy adjustment can trigger widespread service disruptions. Traditionally, change risk assessment relies on manual reviews and experience, with no precise means of validation. Intelligent agents combined with digital twin technology provide an effective solution.

Predictability transforms operations from reactive problem resolution to proactive prevention, with digital twins serving as critical infrastructure. By constructing a virtual network that is structurally identical, behaviorally consistent, and synchronized in real time with the physical network, digital twins provide a zero-risk validation environment for agents. In network change pre-validation, risks shift from post-event detection to pre-event prevention. In fault prediction and early warning scenarios, O&M shifts from reactive response to proactive avoidance. By simulating complex scenarios, the system can evaluate multiple response strategies, select the optimal path, and improve both the quality and

efficiency of handling complex situations.

Paradigm Shift Across Three High-Value Scenarios

When easy interaction, self-optimization, and predictability are integrated into ZTE's OpsAgentStudio core network O&M system, the capability boundaries of the three high-value scenarios—fault handling, complaint handling, and network change—are significantly expanded.

Fault Handling: From Manual Analysis to Agent-Driven Closed Loop

Traditional fault handling faces alarm storms, cross-domain fault isolation, and multi-vendor equipment differences. With the introduction of a fault agent in ZTE's OpsAgentStudio intelligent O&M system, fault handling evolves into an automated closed-loop process. The agent monitors network status in real time, and when a fault occurs, performs alarm suppression, root cause positioning, solution matching, execution, and effect verification. Through self-optimization, it continuously improves its accuracy in handling repetitive faults; through predictability, it identifies early warning signals and intervene before faults occur. Deployment data shows that average fault localization time has been reduced from tens of minutes to within minutes, with some scenarios achieving zero-touch closed-loop handling.

Complaint Handling: From Passive Response to Proactive Prediction

Complaint handling has long been challenged by difficult issue localization, lengthy cycles, and poor user experience. With the introduction of the agent, the complaint handling model undergoes a fundamental shift. Through natural language interaction, users can submit complaint information, and OpsAgentStudio automatically parses the input and extracts key information such as signaling and CHR call records for analysis. At the self-optimization level, the agent correlates multi-domain data, including configurations, KPIs, and operation logs, to accurately identify fault ownership. At the predictability level, it combines historical complaint data with real-time network KPIs to establish a complaint prediction model—when network

quality degrades beyond a threshold, it automatically identifies potentially affected user groups and triggers proactive optimization or notification, transforming complaint handling from reactive post-event response to proactive pre-event intervention.

Network Change: From Window Constraints to On-Demand Changes

Network changes represent the highest-risk scenario in core network O&M. The traditional model relies on midnight cutovers, manual execution, and multi-round verification, resulting in low efficiency and limited business agility. By leveraging the predictability enabled by agents and digital twins, OpsAgentStudio restructures the change model. In scenarios such as disaster recovery operations, configuration changes, and service provisioning, simulation-based verification enables issues to be identified and addressed early, preventing hidden risks from occurring in the production environment. The change process shifts from labor-intensive to agent-led, with change windows expanding from fixed midnight maintenance periods to on-demand daytime execution, significantly improving success rates and reducing manual effort.

Continuous Evolution Toward Autonomous O&M for Future Networks

The introduction of O&M agents is propelling core network O&M into a new phase. "Easy interaction" shifts O&M from command-based to intent-driven operations, reducing collaboration costs. "Self-optimization" enables agents to evolve continuously and autonomously, turning experience into system assets. "Predictability," leveraging digital twins, allows changes and associated risks to be identified and addressed in a virtual space beforehand. Together, these capabilities form the foundation for L4 advanced autonomy.

Looking ahead, advances in technologies such as multi-agent collaboration, real-time cyber-physical interaction, and the integration of large models with O&M knowledge will drive core network O&M agents from assisting humans to taking autonomous control, enabling a high level of autonomy featuring zero touch, zero waiting, and zero failures. In this paradigm shift, agents are no longer merely tools but the core engine of fully autonomous network O&M. **ZTE TECHNOLOGIES**

Integrated Compute-Network-Intelligence Infrastructure for 6G

In the 6G era, intelligence has become a defining feature of infrastructure, driving two trends: AI for Infrastructure and Infrastructure for AI. AI for Infrastructure leverages AI to enable agent-based O&M closed loops, operational optimization, and intelligent orchestration, supporting new O&M models such as digital twins and digital employees. Infrastructure for AI provides ubiquitous, low-latency, and highly reliable compute-network-intelligence resources for emerging scenarios, including embodied AI, connected intelligent vehicles, the low-altitude economy, and XR devices, making computing efficiency and elastic agility critical to compute deployment.

Correspondingly, compute-network infrastructure must meet three goals: full-stack intelligence, network-wide agility, and ubiquitous deployment. Full-stack intelligence requires an integrated foundation for general-purpose and intelligent computing, opening up model inference and AI agent capabilities, and establishing a comprehensive intelligent O&M system. Network-wide agility is enabled by telecom cloud-native architectures, leveraging bare-metal containers for fine-grained scheduling and second-level elasticity, while supporting canary upgrades through stateless architectures and microservices to ensure service reliability. Ubiquitous deployment promotes edge computing deployment and edge collaboration, enabling on-demand orchestration across devices, edges, and the cloud.

Achieving these goals requires technological progress in integrated general-purpose and intelligent computing, cloud-native architecture, and intelligentization. The overall system architecture is shown in [Fig. 1](#).

In the integration of general-purpose and intelligent computing, we first build a unified computing foundation that enables unified

resource pooling, scheduling, and O&M, supports collaboration among heterogeneous hardware such as CPUs, GPUs, NPUs, and DPUs, and evolves toward high-density, high-efficiency, and data-centric computing. An integrated storage architecture supporting block, file, and object storage is developed, adopting an all-flash design optimized for AI training and inference scenarios, with fine-grained hot/cold data tiering and KV cache storage. A DPU-based unified computing substrate, enables unified hosting of virtual machines, containers, and bare-metal servers, while offloading network, storage, and management functions to enhance performance and isolation. Finally, we advance AI compute resource pooling through vGPU technology, enabling fine-grained partitioning, remote invocation, cross-server aggregation, and dynamic scaling to improve resource utilization.

For the practical deployment of integrated general-purpose and intelligent computing, central clouds typically build on existing general-purpose computing pools by expanding general-purpose hardware, deploying new intelligent computing hardware, and upgrading software platforms. This includes reusing existing data center infrastructure and networks, while adding inference computing resources, SSD storage, and lossless RoCE networking. Upgrades to cloud platforms and cloud management enable unified orchestration and management of general-purpose and intelligent computing resources, providing pooled general-purpose computing, GPU/vGPU, and SSD resources to support the diverse computing requirements of AI Core. In edge cloud scenarios, existing integrated edge cloud UPFs can either be upgraded or new ones deployed to provide intelligent edge computing services for terminal agent applications, enabling unified model distribution and agent lifecycle management



Zhu Kun

Chief Engineer of
Cloud Computing
Product Planning, ZTE

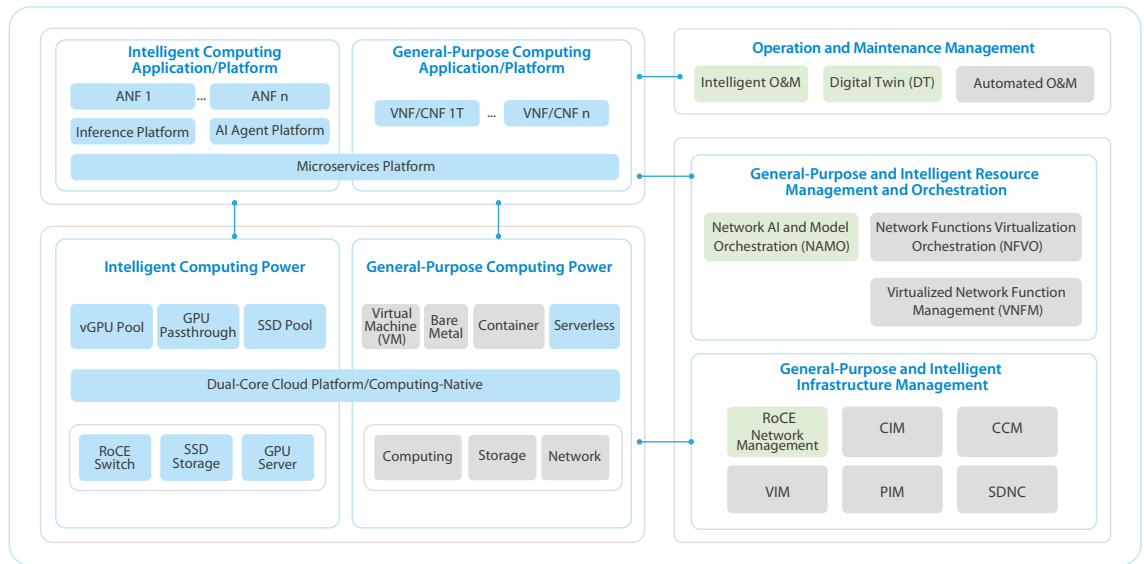


Fig. 1 Integrated compute-network-intelligence Infrastructure for 6G.

through cloud-edge collaboration. Edge clouds are recommended to deploy intelligent computing all-in-one appliances for rapid provisioning, lightweight operation, and enhanced security.

In the cloud-native transformation of network functions, bare-metal containers have become a primary deployment model, offering lower overhead, higher performance, and faster upgrade iterations than traditional virtual machines. The underlying platform needs to enhance dynamic orchestration, large-scale cluster management, and carrier-grade observability. The network function architecture is redesigned based on microservices, using Operators and Helm for lifecycle management and supporting horizontal scaling. To address security challenges like management plane attacks, container escapes, storage and network risks, and OS failure propagation, security is strengthened through security domain isolation, escape detection, DPU offloading, and anti-affinity scheduling.

For network intelligence, a Model-as-a-Service (MaaS) platform provides model fine-tuning, inference, and AI agent services for network operations, maintenance, and service innovation. An agent-centric network deploys operations, O&M, and business agents to enable autonomous perception, decision-making, and optimization, advancing the network toward intent-driven operations with minimal human intervention. An integrated compute-network-intelligence orchestration system

coordinates agents, model services, and traditional network functions, enabling natural-language interaction and dynamic collaboration through large models and agents. A compute-network digital twin provides high-fidelity virtual representations, holographic visualization, simulation, prediction, and decision validation to support autonomous agent optimization.

6G integrated compute-network-intelligence infrastructure faces standardization challenges. In response, a unified architectural standards framework should be established to strengthen cross-organizational and open-source collaboration. An agile, iterative standards development model should also be established to enable the rapid release of phased specifications, while intrinsic security and Zero Trust standards should be advanced to shift from reactive defense to proactive governance.

Looking ahead, integrated compute-network-intelligence infrastructure will become a cornerstone of 6G development. By integrating general-purpose and intelligent computing, advancing cloud-native architectures, and embedding intelligence across all domains, it will overcome NFV/SDN limitations, enable ubiquitous AI, industry digitalization, and AI agent interconnection, and provide a high-performance, low-cost, and autonomous infrastructure for the digital economy and the new era of 6G intelligent connectivity. **ZTE TECHNOLOGIES**

AI+ Signaling Network: Efficient and Reliable Signaling Hub

As the hub of mobile communication networks, the signaling network performs critical control-plane functions, including routing, aggregation, forwarding, and security protection, directly impacting the stability and security of mobile core networks. With the large-scale commercial deployment of 5G networks based on the service-based architecture (SBA), the requirements for 5G signaling networks (NRF/SCP) have grown significantly. For instance, the NRF must accommodate dynamic changes in the 5G network topology. Compared with traditional 2G/3G/4G signaling networks, the number of network functions (NFs) connected to the NRF and SCP has increased by approximately four to five times, making 5G signaling network reliability even more critical.

Two Major Challenges in Routing and Signaling Forwarding

As 5G networks continue to scale, signaling networks face two major challenges: routing and signaling forwarding. First, the explosive growth in NRF service registration and discovery parameter combinations, combined with dynamic topology, leads to routing uncertainty.

- **NF type and service diversity:** 5G defines dozens of NF types, each providing multiple services with different registration and discovery requirements.
- **Multi-dimensional parameters and dynamic topology:** Service registration and request involve multi-dimensional parameters, resulting in massive combinations. Meanwhile, NF instances are distributed across heterogeneous data centers and resource pools, with scaling, deployment, and upgrades continuously changing network topology.

Moreover, 5G networks and NFs experience significant load fluctuations. Traditional rule-based load prediction and flow control have limitations, and SCP-based signaling faces challenges of uneven load distribution.

- **Dynamic load fluctuations:** Traffic tides, user mobility, and heterogeneous hardware capabilities across resource pools cause continuous NF load variations.
- **Limited prediction capability:** Static threshold-based signaling flow control cannot proactively handle congestion or signaling storms.
- **Uneven resource utilization:** Uneven resources and loads across NFs reduce overall resource utilization and resilience against signaling surges.

In summary, NRF must promptly detect abnormal routing outcomes under ever-changing network topologies and massive combinations of service registration and discovery parameters, while SCP must predict signaling load trends for efficient signaling forwarding and flow control.

Empowering the Signaling Network: AI Reshapes Analysis and Processing

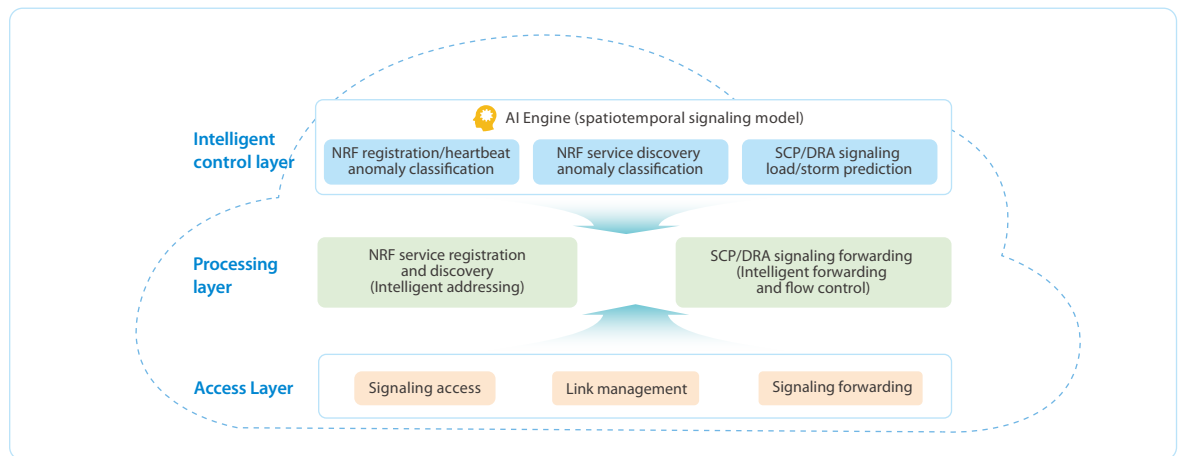
To proactively address reliability challenges in signaling networks, ZTE has launched an intelligent signaling network solution leveraging large AI models. Built on the ZXUN diverse signaling controller (DSC) platform, this solution adds an intelligent control layer, the AI Engine, above the existing access and processing layers. The AI Engine uses a GNN+Transformer framework to build an AI spatiotemporal signaling model, enabling anomaly detection and classification of routing results for NRF service registration, service discovery, and service responses, as well as the prediction of NF loads and signaling storms. The



Zheng Guobin

Chief Engineer of
CCN UDC Product
Planning, ZTE

Fig. 1 AI+ signaling network architecture.



solution architecture is illustrated in Fig. 1.

The ZXUN DSC intelligent signaling network adopts a three-layer software architecture: access layer, processing layer, and intelligent control layer. The access layer handles incoming and outgoing signaling access and forwarding. The processing layer supports NRF-based service registration and discovery for 5GC NFs, while SCP manages signaling forwarding and flow control between 5GC NFs. The newly introduced intelligent control layer provides AI model-based analytical capabilities, including anomaly classification and assessment for NRF registration and discovery results, as well as load trend prediction and policy generation for SCP.

The AI spatiotemporal signaling model adopts a GNN+Transformer architecture and is trained with spatial network topology data and signaling traffic time-series data. The GNN extracts spatial features from network topology data, while the Transformer performs self-attention-based feature extraction from signaling load time-series data.

ZTE's intelligent signaling network provides real-time evaluation and prediction of signaling routing outcomes, enabling signaling load forecasting and control strategy generation.

- **Intelligent identification of known and unknown routing/addressing anomalies:** For each NRF registration or discovery operation, the intelligent signaling network invokes an AI spatiotemporal signaling model to analyze the corresponding results. Abnormal or suspicious results trigger alarms for initial manual

intervention. In the next phase, digital twin-based validation will support intelligent error correction of route addressing outcomes. Laboratory tests demonstrate that the detection rate of abnormal NRF route addressing exceeds 90%.

- **Signaling load prediction and scaling strategy generation:** SCP collects real-time signaling load data from interconnected NFs and invokes the AI spatiotemporal signaling model to predict load trends. When overload or signaling storms are predicted, alarms are issued for manual intervention. The model can further generate scaling strategies, which are applied to the live network after digital twin validation.
- **Built-in AI model with online fine-tuning:** The built-in AI spatiotemporal signaling model eliminates the need for external intelligent processing NFs. It leverages CPU-based inference, online data annotation, and model fine-tuning to continuously improve accuracy, meeting the real-time decision-making and deterministic requirements. The model remains transparent to surrounding NFs and has minimal network impact.

ZTE's signaling products (ZXUN DSC) have been deployed in 210 commercial projects worldwide, covering the signaling network elements across 2G, 3G, 4G, 5G, and IMS networks, and serving over 900 million 4G users and over 500 million 5G users.

With increasing network complexity and the evolution toward 6G, ZTE's AI-native signaling network solution will enable a more reliable and efficient signaling hub, addressing the complexity and reliability challenges of 6G network. **ZTE TECHNOLOGIES**

ZTE Launches Complaint Analysis Agent to Redefine the International Roaming Service Experience

During the 2026 New Year holiday, a tourist from Germany was recording a short video on her phone in Xuzhou when her mobile connection suddenly dropped. She immediately contacted China Mobile International's roaming service hotline to report the issue. To her surprise, in less than 10 minutes, she received an automatic notification stating, "An issue with your network connection has been detected and automatically resolved." The following day, she received a follow-up message: "Thank you for choosing China Roaming. Your feedback will help us further optimize the network."

This scenario exemplifies China Mobile's "Seamless Roaming" initiative in action. In 2025, as inbound tourism fully recovered, the daily volume of international roaming users in China exceeded 3.2 million, while complaints surged by 67% year-on-year. The traditional manual analysis approach faced significant challenges: resolving a single international

roaming complaint required coordination across seven specialized domains—including the wireless network, the core network (CN), and international gateways—resulting in an average resolution time of three hours. With expert resources in short supply, user wait times increased, and customer satisfaction remained under pressure.

Against this backdrop, Jiangsu Mobile, in collaboration with ZTE, launched the first intelligent agent for international roaming complaint analysis in December 2025, transforming the traditional complaint resolution paradigm by leveraging AI. The solution adopts a T1+T2 dual-agent architecture, enabling end-to-end closed-loop automation (see Fig. 1). The T1 agent, operated by the carrier, performs semantic recognition and preliminary classification, automatically generating responses for high-frequency, standardized issues. The T2 agent, powered by ZTE's Nebula multimodal large model, handles complex analytical tasks. When a complaint



Zhang Maopeng

Chief Engineer of CCN
Intelligent O&M
Product Planning, ZTE

involves intricate signaling anomalies, the system automatically retrieves raw signaling and call history records (CHR) without human intervention, reducing expert-level analysis time from several hours to less than 10 minutes.

The technological breakthrough is built upon three key components: a signaling encoding model that precisely parses protocol message byte structures, a projection model that achieves cross-protocol semantic alignment, and a domain-specific decoding model incorporating 3GPP protocol logic. This enables, for the first time, end-to-end signaling inference across more than 100 network elements (NEs) spanning the 5GC, EPC, and IMS without relying on manual labeling. Both accuracy and coverage exceed 90%.

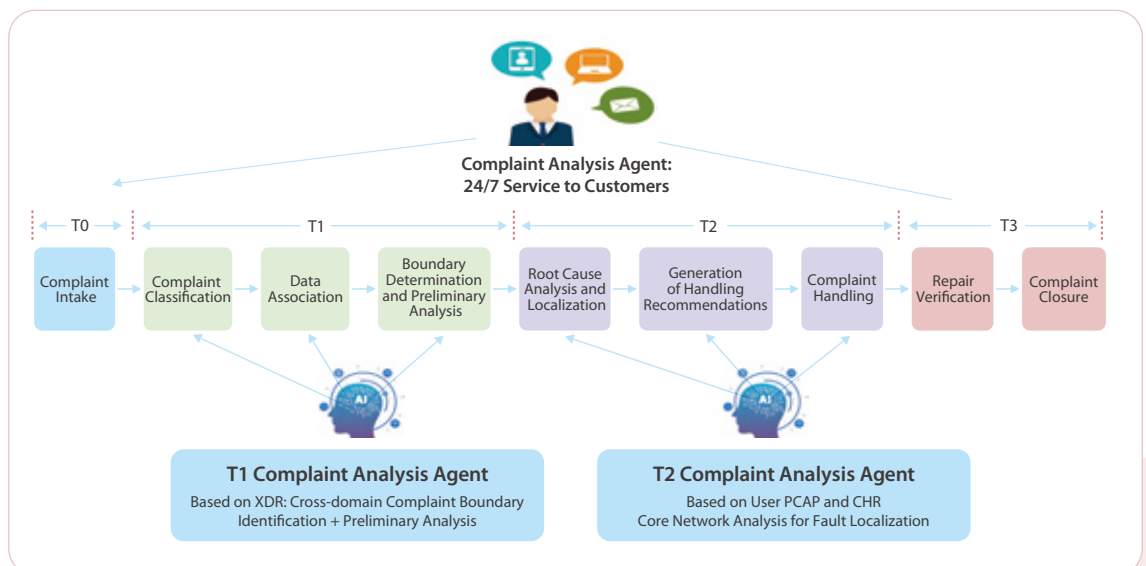
A distinguishing feature of the agent is its autonomous execution capability. Rather than passively responding to directives, it autonomously plans the analysis workflow based on ZTE's proprietary agent reasoning framework, covering intent recognition, signaling swimlane generation, anomaly detection, root cause inference, and case recommendation—all without human intervention. The system dynamically invokes network monitoring tools via APIs, achieving seamless mapping from natural language to executable commands. Operators simply enter queries such as "a user experiencing slow 5G

internet in Nanjing," and the agent automatically correlates location, time period, signaling flows, and QoS parameters to generate a verifiable diagnostic report.

ZTE's complaint analysis agent enables intelligent diagnostics and closed-loop operations for network issues experienced by international roaming users through coordinated breakthroughs across three core technological layers—data, model, and business—significantly improving diagnostic efficiency and user satisfaction.

- **Data layer:** ZTE pioneers a multi-data fusion engine that collects heterogeneous data from multiple sources—including raw signaling from 4G/5G/IMS core networks, CHR, user context, network element configurations, and historical complaints—precisely resolving data sourcing challenges in both data and voice domains under international roaming scenarios and eliminating the "manual data stitching" bottleneck.
- **Model layer:** ZTE has independently developed the Nebula multimodal signaling large model, breaking away from the traditional "signaling-to-text" paradigm. It pioneers native signaling parsing technology to directly process raw byte streams, integrates an enhanced Transformer with multi-level attention mechanisms to capture the semantics and temporal patterns of over 100 network element

Fig. 1 "T1+T2" dual-agent architecture for complaint analysis.





interface protocols, constructs mappings in a unified signaling semantic space, and enables automatic cross-protocol feature alignment. Embedded with a 3GPP protocol knowledge base and a Chain-of-Thought (CoT) reasoning engine, it can autonomously identify root causes such as authentication timeouts, N4 interface anomalies, and gateway configuration errors, with a transparent and traceable reasoning process.

- **Business layer:** ZTE has developed an AI agent for handling core network complaints, implementing an "Observe-Think-Plan-Act-Learn" closed loop. Operators need only ask natural-language questions—such as "Find the cause of slow internet experienced by Chinese users in Malaysia"—and the system automatically invokes APIs, analyzes signaling flows, compares quality metrics, and generates root cause reports with resolution recommendations. With a closed-loop resolution rate exceeding 95%, the system continuously optimizes decision pathways through feedback mechanisms. The toolchain supports automatic API matching through semantic embedding, enabling plug-and-play functionality.

From large-scale deployment in December 2025

to March 2026, the system handled over 100 international roaming complaint cases, reducing the average handling time from 180 to 30 minutes—an 85% improvement—while customer satisfaction reached 98.7% and duplicate complaints dropped by 76%.

During follow-ups, many international roaming users reported that previously, network issues encountered abroad often required multiple contacts with customer support and lengthy waiting times for manual resolution. After the deployment of the system, however, users reported that issues were typically identified and resolved automatically within minutes, eliminating the need for repeated reporting. The system also proactively sends service restoration alerts and solicits user feedback.

This "proactive care" service model significantly improves user satisfaction, contributing to a better international roaming experience. It shifts from people finding problems to AI finding problems, and from reactive response to proactive care.

In the future, the agent will be extended beyond international roaming to cover broader scenarios—including ToC individual users, ToB enterprise dedicated lines, and edge networks—building an AI-agent-powered digital workforce for end-to-end complaint management across domains. **ZTE TECHNOLOGIES**

Network Operations Agent (NOA) V1.0: Enabling a New Paradigm of Intelligent Connectivity



Yuan Xincheng

CCN Solution Director,
ZTE



Wei Beibei

CCN Product Planning
Engineer, ZTE



Chen Xiaofeng

CCN Product Planning
Engineer, ZTE

On December 23, 2025, China Mobile Research Institute and Anhui Mobile, in collaboration with ZTE, successfully completed the end-to-end closed-loop validation of the intent-driven network operations agent (NOA) V1.0 in real-world service scenarios and held its launch event. Taking live streaming service assurance as a pilot use case, the validation showcased a new paradigm of intelligent connectivity featuring user intent-driven control, autonomous network response, and closed-loop capability optimization, marking a critical step in the evolution of operator networks from passive response to proactive service.

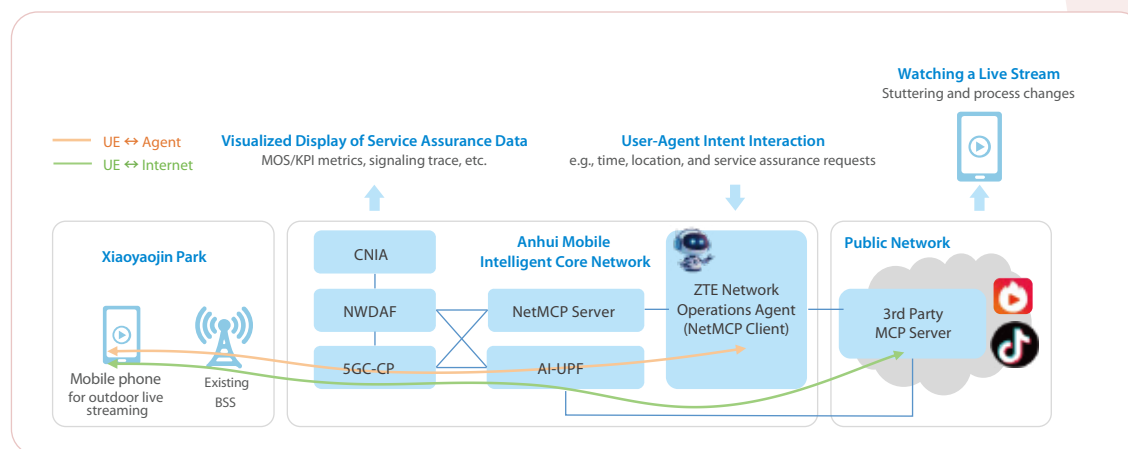
With the widespread coverage of 5G networks and the rapid development of high-definition live streaming, cloud gaming, telemedicine, and other experience-intensive services, user expectations for networks have evolved from basic connectivity to reliable, intent-aware, and proactive services. Traditional O&M models rely on manual configuration and experiential judgment, resulting in slow response, low efficiency, and limited ability to address complex and dynamic service scenarios.

As a pioneer in China Mobile's intelligent transformation, Anhui Mobile, together with ZTE, launched the network operations agent pilot project, focusing on high-occurrence issues such as live streaming lag. The project introduced the innovative

concept of "one network per user, one policy per moment", enabling networks to exhibit human-like intelligence through proactive perception, autonomous decision-making, and closed-loop optimization, moving towards the vision that proactively serve users.

As shown in Fig. 1, this validation employs ZTE's AI appliance as the execution platform, integrating ZTE's proprietary AI agent engine with the network capability orchestration system to build an integrated intelligent operation architecture encompassing intent understanding, workflow orchestration, and the invocation of atomic NE capabilities. It connects core network elements such as CNIA, NWDAF, UPF, SMF, and PCF, enabling automated assurance from user requests to experience assurance within seconds. Through the carrier-grade network model control protocol (NetMCP) developed by China Mobile and ZTE, the AI agent achieves efficient collaboration with the 5GC network platform, moving beyond the rigid traditional API invocation model and enabling the network to understand intent and execute task autonomously. This system establishes an automated, full-lifecycle closed loop encompassing perception, decision-making, execution, and feedback.

- **Perception layer:** The CNIA system collects QoE metrics from live streams (e.g., stall frequency,



◀ Fig. 1 Intent-driven network operations agent enabling end-to-end closed-loop live streaming assurance.

initial frame latency, bitrate fluctuations) to accurately detect anomalies.

- **Decision-making layer:** The network operations agent receives natural-language instructions from users (e.g., "Ensure the quality of the Douyin live stream at Xiaoyaojin Park") and autonomously generates the optimal assurance strategy by leveraging user profiles, service trends, and network congestion predictions provided by NWDAF.
- **Execution layer:** The network operations agent directly invokes 5GC atomic capabilities via the NetMCP protocol, dynamically establishing dedicated UPF bearers, elevating QoS priority, and precisely identifying live-streaming flows to enable on-demand resource orchestration.
- **Feedback layer:** The CNIA system provides real-time feedback on post-optimization QoE metrics. The network post-optimization agent evaluates goal attainment—for example, whether stalls occur fewer than once per minute—and, if the target is not met, automatically refines the strategy to form a self-optimizing closed loop.

With intent-driven network operations agents, users can make natural language requests (e.g., "I want to livestream at Xiaoyaojin"). The system automatically performs intent recognition, risk assessment, policy generation, and resource scheduling. Once the directive is issued, resources take effect rapidly, improving efficiency over traditional manual approaches. Meanwhile, an "Experience Map" is provided to visualize livestreaming-area information, including throughput,

latency, MOS scores, and quality heat maps. In real-world livestreaming trials at Xiaoyaojin Park, stuttering rates dropped markedly, end-to-end average latency was greatly reduced, and network smoothness and responsiveness were enhanced, significantly enhancing user experience.

The network operations agent represents not only a technological breakthrough but also a practical application of Network as a Service (NaaS). It enables on-demand offerings such as the "Livestream Assurance Pack" and the "High-Speed Railway Smooth Streaming Pack," driving the evolution of network interaction from app-based menu selection to natural-language interaction. A single natural-language request triggers an automated assurance workflow, enabling streamlined operations, second-level responses, and improved user convenience and service efficiency, ushering in a new era of AI-driven, user-centric network services.

Looking ahead, ZTE will continue to advance practical technology and open collaboration, focusing on lightweight large-model training and inference, secure agent integration, and coordinated resource orchestration to build an open intelligent network ecosystem. The successful validation of the network operations agent marks a key step toward AI-native telecommunications. Future networks will evolve beyond traditional connectivity pipes into intelligent systems capable of perception, reasoning, proactive service adaptation, and intelligent resource scheduling, creating a new paradigm of "networks that understand you" and supporting digital transformation across industries. **ZTE TECHNOLOGIES**



Partnering with China Mobile to Establish the “Co-Innovation+” Autonomous Network Open Lab



Zhou Bo

Senior System Engineer,
Service Products, ZTE

As global digital and intelligent transformation deepens and telecommunications networks advance toward higher levels of autonomy, traditional O&M models are facing unprecedented challenges. The coexistence of 2G, 4G, and 5G networks and the interconnection of multi-vendor equipment have significantly increased network complexity. Intricate fault propagation paths further hinder efficient root cause localization.

The rise of large language models (LLMs) offers new opportunities for O&M transformation. However, their direct application in highly specialized communication network O&M scenarios still faces several challenges. First, difficulties in multimodal semantic alignment limit their ability to understand network topology and alarm data. Second, AI decision-making remains black-box, with reasoning paths that are difficult to explain or trace. Third, LLMs lack robust causal reasoning capabilities, leading to biased decisions.

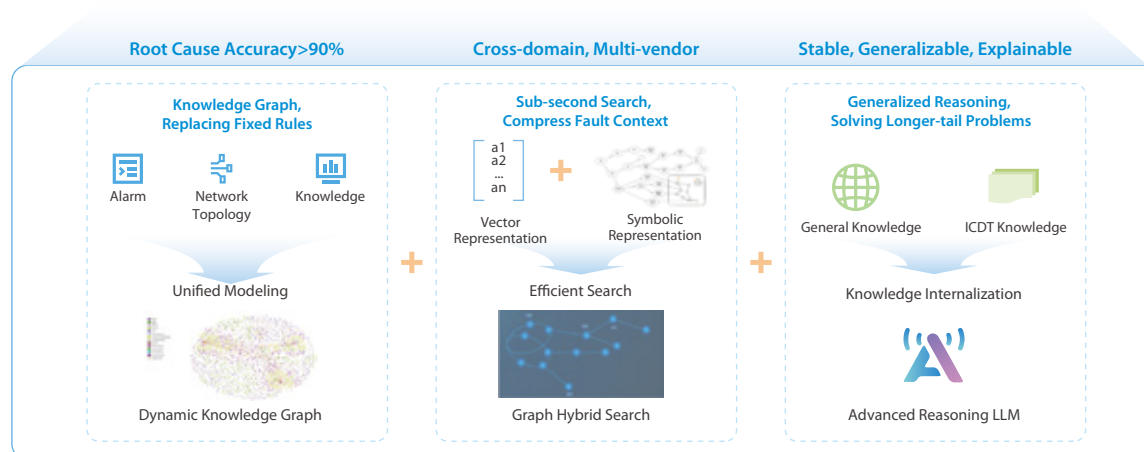
Achieving stable and accurate AI decision-making and shifting from experience-driven to data- and knowledge-driven operation have become urgent industry challenges.

Against this backdrop, ZTE and China Mobile

forged a powerful partnership and officially launched the “Co-Innovation+” Autonomous Network Open Lab in July 2025, focusing on key challenges such as high-precision AI decision-making and multi-agent collaboration. The lab explores a full-chain innovation pathway spanning cutting-edge technology research, live-network pilot verification, large-scale replication, and standards leadership, aiming to break through AI reliability bottlenecks in O&M and empower the high-quality development of autonomous networks worldwide.

Forging a Graph-Based Network Inference Paradigm to Reshape AI Decision-Making

The Co-Lab team proposes a universal solution that empowers fault root cause inference using network graph models (Fig. 1). Rather than simply stacking LLM technologies, it builds an integrated AI decision-making architecture with knowledge graphs as the skeleton, hybrid graph search as the engine, and an advanced reasoning LLM as the brain. Knowledge graphs provide structured network data and knowledge representation, helping the LLM understand operational data and fault propagation mechanisms while constraining



◀ Fig. 1 Graph model based on the deep integration of knowledge graphs and LLMs.

its reasoning scope. This enables LLMs to deliver reliable and accurate root cause localization, accelerating O&M transformation.

- First, a communication-domain knowledge graph is built by integrating static and dynamic data. It realizes unified graph modeling for 11 categories of network resources, including wireless, transport, and power and environmental systems, together with dynamic alarms and fault propagation knowledge. The resource topology graph outlines the logical framework of the network; the fault propagation graph extracts RDF fault knowledge triples from manuals, historical work orders, and expert documents via the LLM; and the dynamic root cause inference graph integrates real-time alarm information to provide AI with up-to-date network facts.
- Second, a hybrid graph search method is introduced to efficiently compress fault contexts. Faced with massive alarms and topology data, it uses spatiotemporal correlations and topological path constraints for subgraph pruning and hierarchical compression. This enables accurate extraction of root cause inference subgraphs from tens of millions of nodes, providing more focused context for the LLM. It reduces token consumption by 30% while improving reasoning efficiency and accuracy.
- Third, in-depth reasoning is enabled through causal Chain-of-Thought (CoT). Based on the fault context within the root cause inference subgraph, the CoT guides the LLM to trace faults layer by layer from alarm nodes along topology links, verify causal relationships through the fault

propagation graph, and perform global reasoning on cascading faults across base stations and domains. In multi-factor outage scenario, the graph model can accurately identify the final root cause, raising the root cause analysis accuracy above 90%.

- Fourth, an end-to-end white-box interpretable decision system is built. The graph model overcomes the black-box limitation of AI by visualizing the entire reasoning process. Alongside analysis conclusions, the system presents the complete reasoning path from fault symptoms through propagation paths to the root cause, while marking referenced fault propagation knowledge entries. This allows O&M personnel to understand the AI's reasoning and significantly enhances trust in its decisions.

Driving Autonomous Network Innovation

These innovative achievements have been validated through pilots across five regions, including Beijing, Guangzhou, Shandong, Hunan, and Jiangsu, covering over 230,000 base stations. The solution demonstrates cross-domain, multi-vendor fault root cause localization capabilities, with an accuracy of over 90%.

Meanwhile, the lab team has jointly produced a series of high-value patents and papers in the network graph model field, built the first graph model evaluation dataset for the telecom domain, and conducted research on the Agent-to-Agent for Telecom (A2A-T) protocol specifications. These efforts have created valuable technical assets for autonomous network technology innovation and set a paradigm of open collaboration and joint value creation worldwide. **ZTE TECHNOLOGIES**

ZTE

To lead in connectivity and intelligent computing, enabling
communication and trust everywhere