AUG 2024 | VOL. 26 • NO.4 • ISSUE 213

ZTE TECHNOLOGIES

VIP Voice

Ncell: Empowering Nepal to Go Digital

Executive Insights

ZTE CEO Xu Ziyang: Ingenuity for Solid Foundation, Openness for Win-Win

Special Topic

AI-Enabled Network Services

Scan for mobile reading

Cover Figure | *Jabbor Kayumov, CEO & Managing Director of Ncell*

ZTE TECHNOLOGIES

AUG 2024 | VOL. 26 • NO. 4 • ISSUE 213

Advisory Committee

Director: Liu Jian
Deputy Directors: Sun Fangping, Yu Yifang, Zhang Wanchun, Zhu Yongxing
Advisors: Bai Gang, Fang Hui, Hu Junjie, Hua Xinhai, Kan Jie, Li Weizheng, Liu Mingming, Lu Ping, Tang Xue, Wang Quan, Zhang Weiqing, Zheng Peng

Editorial Board

Director: Lin Xiaodong
Deputy Director: Huang Xinming
Members: Deng Zhifeng, Dai Yanbin, Huang Xinming, Jiang Yonghu, Ke Wen, Kong Jianhua, Liang Dapeng, Liu Shuang, Lin Xiaodong, Ma Xiaosong, Shi Jun, Xia Zejin, Yang Zhaojiang, Zhu Jianjun

Sponsor: ZTE Corporation

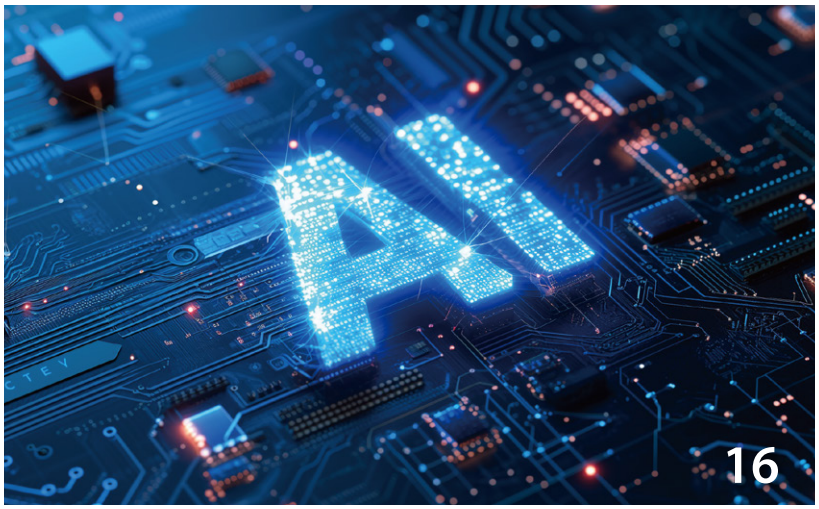
Edited By Shenzhen Editorial Office
General Editor: Lin Xiaodong
Deputy General Editor: Huang Xinming
Editor-in-Chief: Liu Yang
Executive Editor-in-Chief: Yue Lihua
Special Topic Editors: Zheng Hong, He Fang, Yang Yan
Circulation Manager: Wang Pingping

Editorial Office

Address: NO. 55, Hi-tech Road South, Shenzhen, P.R. China
Postcode: 518075
Website: wwwen.zte.com.cn/en/about/publications
Email: yue.lihua@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.

CONTENTS



Success Story

Ncell: Empowering Nepal to Go Digital

Reporter: Luo Jinfeng

Jabbor Kayumov
CEO & Managing Director of Ncell

Ncell, the first private mobile service provider in Nepal since 2004, has played a significant role in bridging the country's connectivity gaps. However, it now faces the challenging task of accelerating digitalization in Nepal, where 4G capacity is still not fully utilized. "We, as enablers and infrastructure providers, have a responsibility to support mass consumers and various sectors towards this digital future," says Jabbor Kayumov, CEO & Managing Director of Ncell. He discusses the challenges confronting Nepal's telecom sector, emphasizes the importance of all stakeholders working effectively to shape a digital future, and shares Ncell's plans to addresss customers' digital needs.

As a landlocked country, Nepal faces challenges in developing its telecom infrastructure. What are your thoughts on the country's evolving telecom landscape and its potential for growth?

Being landlocked is a challenge, but it can also be seen as an opportunity with two of the fastest growing large economies of the world beside us, connected by the information super highway. Recognizing the vast potential of this country, we embarked on a journey to bridge Nepal's connectivity gaps as the first private sector GSM mobile operator in the country nearly two decades ago. Since then, the telecommunications sector has witnessed a paradigm shift, and our partners, like ZTE, have been supportive in this journey.

Topographical diversity was indeed a significant challenge when we had to reach far-flung areas to serve our customers. There were many instances where we had to have heavy tower equipment hand carried or use donkeys and yaks, and even airlift materials to install towers. Due to the lack of road accessibility, such requirements significantly increased the cost of infrastructure development. Additionally, the need for permissions from multiple agencies for setting up a tower and community issues added to the challenges.

Our journey in Nepal has been quite a feat nevertheless. Today, Ncell stands as a proud service provider connecting over 13 million customers nationwide, ensuring access to basic telecommunication services, as well as modern high-speed 4G connectivity to over 92 percent of the country's population.

Despite the nationwide expansion of 4G, Nepali consumers are not fully utilizing the country's 4G capacity, however. This presents both a challenge and an opportunity to fuel digitalization so that more and more people tap into the 4G opportunity and consume more data to go digital. The future is digital, and we also need to transform from a telco to a tech-co. We have started working towards this goal.

How do you envision Ncell advancing under your leadership, and what challenges are you currently facing?

There is no option but to go digital, and I am sure that is also why the government of Nepal has embarked on its Digital Nepal Framework. Yet the pace of digitalization is relatively slow here, and we, as enablers and infrastructure providers, have a responsibility to support mass consumers and various sectors towards this digital future. From 2G to 3G to 4G, the market is now moving towards 5G. Yet with 4G still underutilized and telecom operators going through a tough phase, the question arises: are we ready for 5G? Industry data shows the telecom sector's revenue has declined by more than 25 percent in the last seven years. Despite some growth in data consumption, the revenue increment from data is nominal and does not compensate for the decline resulting from a fall in voice and ILD revenue. Our average monthly data consumption per customer per month stood at 177 MB in 2017, which has now increased to over 4 GB. Yet, if you look at income from data, it has only increased by NPR 1 billion in six years to Rs. 12 billion. People are more attached to using OTT services using WiFi or mobile data, which has been hitting both operators—Ncell and Nepal Telecom—hard.

If we envision a 5G future in Nepal, it needs be commercially viable, and this will only happen when customers use more data, there are more digital services, and an improved ICT ecosystem. 5G is about enterprise, business, Internet of Things, and machine-to-machine communication where everything is connected. They might have physical SIM cards or e-SIM solutions, and all these machines have to be connected through operators like us. Such solutions might come from Ncell or other third-party service providers. In order to create a market for connected devices, 5G, or a digital future, all stakeholders—government, service providers, customers, and other components of the ecosystem—have to work effectively to create value for each other.

For example, the government has to come up with favorable policies on spectrum allocation, service taxation, renewal of licenses and fees, equal level playing field, and service expansion, encouraging operators to invest and contribute to digital development.

If these issues remain unaddressed, it will be

difficult for telecom operators to sustain in the market. There is also a big gap in content in the market. When we talk about digitalization, there needs to be maximum use of content and applications in diverse sectors such as public service delivery, education, health, and agriculture to name a few, which ultimately contribute to digital evolution.

What key efforts are you implementing to tackle challenges and reinforce your position in the country?

Our market has changed significantly, and so has our focus on data services. The majority of our CAPEX goes to data services, yet the reality of the returns I have already shared. Ncell has become a big brand and our consumers have expectations from us. We are always committed to addressing the demands of our customers, who made Ncell what we are today. Being a Nepali company, it is our responsibility to serve Nepal and Nepalis, and we are aligned with the theme 'Here for You'. It means we are here to listen and provide solutions that consumers need. We plan to come up with solutions that lead our customers to go digital and set a pathway for 5G and beyond.

At Ncell, we have more than 98% Nepali colleagues equipped with good knowledge, enabling Ncell to be a world-class service provider. We have a vision to make Ncell a complete digital technology company, and from the capability, knowledge, and technical expertise perspectives, we are well-equipped to do the best for the country. We also now have support from our strategic partner—e& international. We are poised with a strong commitment and capability to diversify our digital services, enriching the lives of Nepali people across the country. The imperative for ongoing expansion of network infrastructure, beyond mobile connectivity, further underscores the significant opportunities that lie ahead for the country.

What's your strategy going forward?

There are markets like our neighboring countries Sri Lanka and Bangladesh, which are doing very well in terms of revenue. The telecommunications industry has devalued itself from the value-add perspective. Today, a lunch set of a Thakali meal costs more than the monthly phone bill. Here, consumers can simply compare

the value added. My point is that the value that our connectivity gives is actually significant, but there is still a perception that it is expensive. What is more important is that we have never increased the price of our services no matter what the inflation rate is and how much we have invested to make our network available nationwide.

As a service provider, we need to survive and continue bringing the latest technologies and the best quality of service. Today, the growth will come from non-traditional solutions that address consumer needs in multiple areas like education, health, public service delivery, and entertainment, etc., through the use of the internet. For this evolution, service providers like Ncell are going to be key for future growth. The role of the government, or say the regulator, is also very critical because it is the responsibility of the regulator to make sure that industries and companies like us will have the returns to ensure continued investments in the latest technologies as this industry is very capital-intensive.

With many markets around the world already embarking on their journey to 5G, there is also an expectation that Ncell will introduce 5G in the Nepali market. Yet, to bring this new technology, operators need to make a huge investment. At a time when the telecommunications industry is already facing problems of continuous revenue decline, the regulator needs to ensure that the cost of the spectrum or licenses is set ensuring that service providers earn profits and in return, the companies will continue to invest in new technologies. Telecommunications is a sector where you need to build, maintain and keep reinvesting to introduce newer technologies.

Collaboration is key to success in the telecom industry. How do you collaborate with your partners to drive growth and innovation? Having cooperated with ZTE in wireless network projects, how do you evaluate ZTE's competence in project delivery and provision of high-quality products and services?

The telecommunications sector has always been a driver, or you could say an enabler, for every other sector. With changes in technologies, in which it rides and shifts towards broadband and beyond, many things have changed in the last decade. Many things we do are paperless today; even many services are now being provided electronically, and after the COVID-19 pandemic, consumer behavior has shifted more towards data. Following the launch of 4G, the cost of mobile broadband has come down drastically, contributing to the economy and encouraging consumers to gradually go digital. These days, we see almost all businesses using scan codes for digital transactions. We are happy to be a major contributor to this evolution that is taking place. There still lies huge potential in the digital sphere when it comes to solutions that make the daily lives of general consumers and the performance of businesses better and more productive.

Where we stand today is a result of industry collaboration. As our market is still behind other economies in terms of digitalization and reaping benefits from it, there are more things that we can do to diversify the use of data. For that, we need the support of the government, and content creators need to work on applications that best suit the Nepali market and help consumers adopt a digital life. What I find here is that our market has low digital literacy, which results in lower usage of applications and data consumption. Our focus is to drive data, change consumer lives, and contribute to the economy's digitalization. For that, all stakeholders, including partners like ZTE, have a big role to play.

In our journey, we have collaborated with ZTE for a long time on multiple wireless network projects, including radio/baseband, core network, transport network service, and network operation and maintenance. We believe that ZTE will focus more on updating available technical solutions which bring simplicity and efficiency in managing operations. Also, we are pretty confident that together with ZTE's technologies and solutions, we will continue to create good value for our customers and the country while moving ahead with the digital development of the country. ZTE TECHNOLOGIES

ZTE CEO Xu Ziyang: Ingenuity for Solid Foundation, Openness for Win-Win

Editor's Note: ZTE CEO Xu Ziyang has delivered a keynote speech at the "AI First" session at MWC Shanghai 2024. He shared ZTE's practices and innovations in intelligent digitalization amidst the AI wave.

Embracing Changes and Promoting Intelligent Evolution

Over the past year, large language models and generative AI have accelerated the transformation towards an increasingly digital and intelligent world. With the rapid emergence of new technologies and products, new business scenarios and models are also gaining momentum. Although generative AI is still in the early stage, there is a growing consensus that the world has already entered an AI-driven industrial revolution. AI will have disruptive and far-reaching impacts on all aspects of production and life, and significantly reshape the global economic landscape. According to forecasts by a consulting firm, by 2030, AI will boost China's GDP by 26% and North America's by 14.5%. This is equivalent to USD 10.7 trillion and accounts for almost 70% of the global economic impact. Apparently, AI will bring unprecedented business opportunities in various sectors such as retail, financial services, and healthcare.

Apart from issues concerning hallucinations, security, and ethics, the development of generative AI also faces challenges in terms of computing power, energy consumption, dataset, standardization, commercial application, etc. Therefore, advancements in multiple areas are required. As such, ZTE proposes three major principles: computing and network evolution, training and inference enhancement, openness and decoupling.

First, to break through technology bottlenecks, it's crucial to strengthen research on architectures, algorithms, computing networks, and hardware-software synergy, thus improving AI training and inference efficiency. Second, various solutions such as retrieval-augmented generation (RAG) and AI agents should be employed to ensure reliability, security, and interpretability, thereby facilitating the widespread application of large models and higher value creation, and building a data flywheel that improves both capabilities and business efficiency. Finally, accelerating standardization through openness and decoupling will help build a thriving industrial and commercial ecosystem.

Building a Highly-Efficient Foundation Through Computing and Network Evolution

To begin with, we emphasize computing and network evolution to build a highly-efficient foundation. For intelligent computing, high-speed network connections are not just vital but also indispensable. From die-to-die (D2D) connectivity to interconnects of chips, servers, and data centers, continuous innovation and breakthroughs in

Xu Ziyang, Chief Executive Officer of ZTE

network technology will greatly enhance the performance and efficiency of intelligent computing.

More specifically, die-to-die emphasizes high-speed interconnects between bare dies in a single package. Combined with the full series of in-house parallel and serial D2D interface IPs, as well as the advanced 2.5D and 3D packaging technologies, our solution enables heterogeneous integration and disaggregation. To a certain extent, challenges arising from the slowdown of Moore's Law and constraints in manufacturing can be effectively mitigated. We have developed chip architectures that enable heterogeneous computing and network processing, which in turn deliver enhanced performance and cost efficiency.

Chip-to-Chip focuses on interconnects across chips, which enables the development of a solution that integrates distributed high-speed interconnect and a full range of interfaces including PCIe 5/6 and 56G/112G/224G SerDes. This solution can effectively address the inflexibility and low bandwidth utilization of the mesh interconnect architecture. Furthermore, with the tensor parallelism degrees of up to 8/16, this solution better adapts to complex, large-scale intelligent computing scenarios, thus providing customers with differentiated competitive edges. By virtue of such a solution, ZTE has played an important role in a major operator's advancement of the OISA architecture.

Furthermore, to meet the requirements of heterogeneous integration of photonic and electronic ICs for the next-gen 102.4T network switches, linear-drive pluggable optics (LPO) and co-packaged optics (CPO) can significantly enhance the interconnect density and reduce power consumption. In addition, optical I/O brings revolutionary improvement in bandwidth density, power efficiency, and latency.

Server-to-Server covers interconnects among intelligent computing clusters. In this regard, for example, ZTE has been fully collaborating with China Mobile to foster a robust industry ecosystem of global scheduling ethernet (GSE), aiming to build a non-blocking, high-bandwidth and ultra-low-latency network for the next-generation intelligent computing centers. In February, we participated in the GSE prototype interoperability test organized by China Mobile. In future scenarios involving thousands or tens of thousands of GPUs, ZTE will work with partners to promote intelligent

innovation, contributing to the development of the industry chain. In terms of capability enhancement, ZTE will continue to improve the forwarding capacity of key chips from 12.8 Tbps to 51.2 Tbps. By offering diverse solutions ranging from single-layer, box+box, to chassis+box, we aim to better serve the needs of AI training in all scenarios.

DC-to-DC falls into the scope of wide-area connectivity. As we know, the 400G OTN lays a solid foundation for the intelligent interconnection of data centers. ZTE has made every effort to facilitate the commercial deployment of the world's largest 400G OTN, and collaborated with operators to demonstrate high-capacity connections of Real 400G. Together, we have completed the industry's first real-time Tbit-level transmission on a single fiber based on S+C+L wavelength bands, setting a world record for transmission distance. Going forward, ZTE will give full play to our strengths in connectivity to build efficient and all-optical networks for intelligent computing.

Empowering Various Sectors with Enhanced AI Training and Inference

For the application of large models across various sectors, in addition to the common issues concerning hallucinations, security, and ethics, it is also necessary to address a series of challenges in critical scenarios, such as expertise, accuracy, robustness, and traceability. Meanwhile, for the building of domain-specific models based on a foundation model, expertise in data governance, incremental training, and related engineering experiences and toolsets are also crucial. Take the telecom autonomous networks as an example, technological innovations such as the integration of large and small AI models, RAG, multi-agent collaboration, digital twins, and multimodal chain of thought (CoT) have all yielded promising results.

In the commercialization of AI application, training and inference are crucial to the leapfrog growth of the real economy. With massive application scenarios and proprietary data, China can make significant contributions to the global AI industry. To fully leverage this advantage, we must focus on enhancing accuracy, expertise, and inference efficiency, while also strengthening domain-specific data governance and the application of digital twin technology. Training boosts AI capabilities, while inference and application bring commercial value. During this circulation of mutual enhancement, a data flywheel is

established. This will further accelerate the improvement and monetization of AI capabilities, which then turn into our core competence.

Then how to achieve that? We emphasize partnerships with high-value industries and angel customers. As the most influential players in an industry, angel customers are usually equipped with robust digital infrastructure. They actively embrace technological transformation and can lead the entire industry in the process of digital and intelligent evolution. By collaborating with these customers, we can not only integrate intelligent technologies with know-how, but also enable rapid validation and refinement of technical solutions to create exemplary use cases. Based on in-house or open-source foundation models, we can develop domain-specific large models with extensive industry data and knowledge engineering, thus making breakthroughs from 0 to 1. And these domain-specific models can be deployed from 1 to N by adapting to different application scenarios.

Openness and Decoupling for a Prosperous Ecosystem

Finally, it's about embracing openness and decoupling to build a prosperous ecosystem. Despite the fast iteration of AI technologies, the current AI ecosystems remain closed, with industry-wide standards yet to be developed. This leads to several problems, such as resource waste caused by redundant development, risk concentration due to technology silos, and supply chain monopolies as a result of limited choices, all of which constrain the rapid and healthy development of AI.

Against such background, ZTE proposes a full-stack, open intelligent computing solution.

At the infrastructure level, hardware-software synergy maximizes resource efficiency. Specifically, the hardware is compatible with mainstream GPUs/CPUs in China and abroad, and supports open standards such as OSIA and RoCE/GSE for high-speed and lossless interconnection, offering customers a variety of choices. The software supports heterogeneous resource management, training and inference job scheduling, and heterogeneous collective communication. Compatible with GPUs from multiple manufacturers, the software enables a high-performance and reliable runtime environment for models. In addition, technologies such as computation offloading and in-network computing significantly reduce data read, write, and transmission time, thereby improving computing utilization rate.

In terms of capability platforms, the solution adapts to mainstream frameworks such as PyTorch and TensorFlow, enabling automatic backend compilation and optimization. Also, it provides an end-to-end engineering toolkit for data processing as well as the development, training, optimization, evaluation, and deployment of models. In addition to full lifecycle assurance and management, the solution also supports compute-native networking, heterogeneous training, efficient inference, and data flywheel building.

As for computing networks, the solution enhances computing and network synergy, facilitating seamless application migration across domains.

As a Chinese saying goes, "A single flower does not make spring, while one hundred flowers in full blossom bring spring to a garden." Similarly, a full-stack, open intelligent computing solution contributes to an open technology ecosystem and a

win-win business ecosystem, which will in turn advance the healthy development of intelligent computing.

Through the decoupling of hardware and software, training and inference, and models, composable capabilities can be developed and widely shared, accelerating the innovation, R&D, application, and commercial use of AI technologies. All of these contribute to an open technology ecosystem.

Through collaboration between chip manufacturers, ICT hardware manufacturers, application developers, integrators, and operators within the industry, we can grow stronger together, building a win-win business ecosystem.

Digital Infrastructure and AI Transformation for a Better Future

Focusing on customer value, ZTE provides a full-stack and full-scenario intelligent computing solution involving computing power, networks, capabilities, intelligence, and applications. Multiple key technologies are also in place, such as high-speed interconnection, in-network computing, compute-native networking, seamless migration, data processing, and algorithm optimization. We focus on building efficient, green, and secure computing networks as digital infrastructure, and apply flexible, agile, and intelligent capabilities and applications for AI transformation. With the fully open composable R&D architecture of ZTE Digital Nebula, we can flexibly collaborate with customers by giving full play to our complementary strengths, thus empowering the digital transformation of industries.

In terms of digital infrastructure, ZTE provides a full series of computing, storage, network, and data center products and solutions, to fulfill various construction needs for intelligent computing centers from the core to the edge.

Regarding computing power, our chips are compatible with GPUs/CPUs from multiple manufacturers in China and abroad. Also, a series of products, including AI servers based on mounted modules/PCIe cards and integrated training and inference cabinets are developed to flexibly adapt to various scenarios. In addition, relying on energy-saving technologies such as hybrid cooling and scalable power distribution, we have developed a new intelligent computing center solution, with a PUE as low as 1.1 and a maximum cabinet power density of up to 60 kW.

In the realm of networking, ZTE, together with industry partners, has proposed an open GPU interconnect standard known as OpenLink (OLink), which will be fully integrated into the OSIA architecture of China Mobile. Compatible with the RDMA protocol for unified intra-node and inter-node communication, OLink focuses on promoting intra-node communication by shifting from mesh interconnects to switch-based interconnects. Such solution takes advantage of tensor parallelism for large-scale computing on a single node, reducing connection complexity, and improving cluster scale and efficiency. In addition, the self-developed RoCE NICs as well as box and chassis RDMA switches support the building of intelligent computing clusters that involve up to 100+ to 10,000+ GPUs. Also, the Real 400G solution helps build efficient and all-optical networks for intelligent computing.

For capability enhancement, we have developed various capability platforms. ZTE TECS, our unified cloud management platform, supports heterogeneous resource management, training and inference job scheduling, and heterogeneous collective communication. Meanwhile, ZTE AIS, a training and inference platform, can be applied to data processing and development of large models, providing full-stack engineering toolkits and engines. Taking seamless migration as an example, availability can be achieved within 5 days, and optimal performance within 15 days. Moreover, efficient inference enables the deployment of trillion-parameter models on a single GPU, and automated data labeling saves 80% of the time.

With regard to intelligence, ZTE Nebula Large Model focuses on algorithm innovation, data engineering, and efficient computing. The foundation model is available in various parameter sizes such as 2.5B, 16B, 40B, and 100B, and can be deployed in mobile, edge, and central cloud scenarios.

In the training phase, innovative technologies are

> **ZTE adheres to the principles of diversity, collaboration, and openness for win-win success, and firmly supports and promotes the prosperous development of industries. Meanwhile, we will double our efforts in intelligent computing, continuously leading innovations and development.**

applied, such as multi-stage pre-training, Chinese vocabulary improvement, high-quality corpus refinement, and synthetic data training. These innovations ensure effective model training while reducing the consumption of computing power by 50%.

As for inference efficiency, by quantizing weights to INT4 and KV cache to FP8, inference resources are saved threefold without compromising model accuracy.

Building on these advancements, the Nebula Telecom Large Model has gone through training based on mixture of experts (MoE) architecture with trillion parameters (9*20B). This model supports multimodal input and a context window of 120,000 tokens, providing expert-level insights and assistance for telecom business scenarios. At the same time, through multi-model collaboration, the Nebula R&D Model can be used in over 30 scenarios at different stages of the entire process, from requirement analysis, design, programming, to testing. Also, it can generate code in multiple programming languages, such as Python, Java, C/C++, Go, and JavaScript, achieving performance on par with GPT-4. Additionally, it significantly surpasses GPT-4 Turbo in terms of the accuracy and coverage of unit testing, and can directly generate test cases based on requirements (test-driven development).

Speaking of applications, ZTE is actively exploring the practical use of large models across industries. ZTE's large models have been applied in various fields such as R&D efficiency improvement, telecom network O&M, urban governance, and industrial parks. In particular, the Nebula Coder Model ranks among the top tier in terms of HumanEval scores. It currently has over 13,000 daily active users, handles more than 110,000 daily requests, and processes up to 330 million tokens per day. This model has improved coding efficiency by 30% and overall R&D efficiency by 10%. Meanwhile, through an end-to-end intelligent computing platform, ZTE provides customers with a whole-process large model toolchain, lowering the barriers to AI adoption and reducing development and usage costs. It can be said that ZTE Nebula Large Model truly brings intelligence to various industries.

To conclude, ZTE adheres to the principles of diversity, collaboration, and openness for win-win success, and firmly supports and promotes the prosperous development of industries. Meanwhile, we will double our efforts in intelligent computing, continuously leading innovations and development. Going forward, ZTE will continue to increase R&D investment and dedicate its efforts to achieving technological leadership in multiple fields. In this way, we hope to further promote the growth of the intelligent computing industry, contributing to economic prosperity. ZTE TECHNOLOGIES

Accelerating Network Service Intelligence to Achieve New Value

Sun Fangping
SVP, President of Global Services of ZTE

In the development process of the telecommunications industry, communications networks have gradually taken on new characteristics such as high complexity, super-large scale, and high dynamics in recent years, influenced by technological innovations, market demands, and policy environments. This poses new challenges to network deployment and O&M.

It is now widely acknowledged in the industry to actively use digitalization and AI technologies to accelerate the transformation of communications networks. According to Omdia, a global analyst and advisory leader, more than three-quarters of enterprises in the communications industry have deployed or trialed AI-based services across multiple areas including customer support, fault diagnosis, and network planning.

As the world's leading provider of integrated communications and information solutions, ZTE actively promotes digital and intelligent transformation in the deployment and operation of communications networks, and focuses on "ultra-efficient deployment, ultimate experience, and network security", helping global operators embrace the era of AI.

Building Digital Infrastructure with Ultra-Efficient Deployment

To improve efficiency, ZTE adopts the solution of "all-domain digital and intelligent collaboration with single-domain intrinsic autonomy". For all-domain digital and intelligent collaboration, ZTE applies digital intelligence technologies across two dimensions: business processes (engineering, technologies, and services) and product networks (wireless, wired, core network, and digital energy). This approach facilitates data collaboration across the whole domain, achieving efficient network deployment and operation. Regarding single-domain intrinsic autonomy, ZTE develops multiple intelligent tools for individual domains of services and products to enhance the capabilities of single modules of network planning, construction, optimization, maintenance, and operation.

All-Domain Digital and Intelligent Collaboration

Continuously promoting digital innovation, ZTE has developed its own digital delivery platform called intelligent engineering project management system (iEPMS). iEMPS features “data & computing

dual-drive", "real-time perception", "intelligent decision-making", "outward extension", "smart collaboration", and "multiparty win-win". This platform covers end-to-end processes including contracts, supply chain, procurement, engineering, services, and finance, helping improve network construction efficiency. Based on the overall five-year plan of digital and intelligent evolution, this platform is progressing towards the intelligent online stage. In 2023, iEPMS achieved interconnection with the systems of operators in provinces and cities across China, including Zhejiang, Sichuan, Henan and Beijing, promoting both network construction efficiency and quality.

In the technical delivery field, ZTE has launched the digital and intelligent technical delivery platform iTech. Different from other models where products and tools are separated from digital services in the industry, the iTech architecture design has a leading edge. By using innovative technologies such as AI and big data, along with digital service capabilities, and supporting multiple products such as wireless, bearer, core network, and big data, the iTech platform builds an ecosystem of intelligent tools to achieve end-to-end service interconnection. The iTech platform provides full-service coverage, including precise network planning, synchronization between survey data and engineering parameters, automatic site commissioning, intelligent acceptance tests, network self-optimization, and intelligent network maintenance. Through the application of the iTech platform, communication time in network deployment is reduced by 50%, commissioning efficiency is increased by over 30%, network-wide faults decrease year-on-year by 8.5%, and emergency fault recovery time is decreased by 20%, greatly improving the network deployment efficiency and network performance.

Single-Domain Intrinsic Autonomy

The single-domain digital intelligence tools developed by ZTE encompass all communications products, including wireless, wired, core network, and digital energy, and cover the whole service process, from planning and construction to optimization, maintenance, and operation.

Taking wireless network planning as an example, ZTE has developed the intelligent planning tool VMAX-RSH, with big data+AI at its core. This tool employs intelligent analysis and comprehensive evaluation of multiple dimensions such as network coverage, capacity, user perception, and network competition to accurately, rapidly, and automatically output 4G/5G network planning solutions with an accuracy of over 80%. In addition, it can implement multiple tasks in parallel, enabling independent output of solutions. In 2023, this platform facilitated intelligent planning for over 50 projects around the world, effectively improving network planning efficiency and accuracy.

In engineering construction, multiple single-domain scenarios, such as iEPMS-based engineering plan management, material management, quality management, and acceptance management, are moving towards intent-based intelligence. For instance, AI review utilizing deep learning and image recognition has replaced manual review in the engineering quality inspection of communications networks. With the help of a scenario-based AI algorithm, quality work orders are immediately inspected upon submission, and immediately reviewed upon inspection, significantly reducing labor costs and shortening review cycles. Statistics show that the average quality inspection period for quality work orders has been shortened from 4.5 days to 20 minutes, improving network construction efficiency and comprehensively guaranteeing network construction quality.

Achieving Value-Driven Business Success with Ultimate Experience

To build future-oriented network O&M, it is necessary to focus on intelligent O&M based on user experience and the exploration of commercial potential. Leveraging the UniSeer intelligent O&M solution alongside AI and big data technologies, ZTE has proposed a "service-centric" active O&M model, aiming to create a mapping between user perception, service quality, and network status.

In terms of user satisfaction assurance, this solution implements all-domain insight, identifies and locates network problems and hidden troubles

that may affect service quality and user perception, and achieves an intelligent closed loop of service experience through end-to-end assurance.

For network O&M, this solution implements intelligent closed-loop O&M functions such as large O&M models, automatic process engines, and event management, injecting intelligence into the O&M process. This transforms offline O&M into online O&M, and provides intelligent optimization for various complicated scenarios.

For business development, this solution supports business operation expansion through digital and intelligent profiles, the digital and intelligent empowerment of edge NaaS, and DaaS+SaaS.

In UniSeer intelligent O&M projects, ZTE gradually promotes the ultimate goal of "three-zero, three-self", providing users with a "zero waiting, zero touch, and zero fault" customer experience from a service perspective, and constructing a new network O&M experience of "self configuration, self healing, and self optimization" from a network perspective. ZTE utilizes deep packet inspection (DPI), digital twins, and intent engines to create a smart brain that enables cloud network integration and intelligent computing for the digital intelligence center, ensuring digital and intelligent empowerment for operators' capability centers and assisting them in reducing costs, increasing efficiency, and transforming from manual to digital O&M. This solution visualizes operators' networks, automates O&M processes, intelligently analyzes technical problems, and enables rapid and remote operations in equipment configuration, O&M process automation, and network optimization. This reduces the impact on users, providing them with a stable, high-quality, and intelligent experience, and supports operators in digital transformation across planning, construction, optimization, maintenance, and operation.

Through the implementation of intelligent O&M, terminal user satisfaction of several operators' networks, which ZTE provides with managed services, has increased by 10%–20%, the number of users has grown by 10%–35%, network traffic has surged by up to 130% annually, resulting in significant improvements in operator-level benefits.

Safeguarding Network Operation Security with AI Technology

With the rapid development and wide application of cloud computing, big data, the Internet of things, and AI technologies, the scale of systems and volume of data are experiencing explosive growth. As network virtualization and the convergence of new and legacy networks continue, the network topology structure becomes increasingly complicated. The diversification of network attack methods also brings unprecedented challenges to network operation security. ZTE adopts a prevention-oriented and

> **ZTE has launched the "Digital Nebula" solution and platform, which enables higher-order intelligent applications in communication network service fields such as intelligent Q&A, intelligent network assistant, intelligent preventive maintenance, precise customer service, and network insight.**

automatic control-oriented security management concept and practice, and incorporates communications large models and AI capabilities to guarantee network operation security.

- **Cloud network:** A heterogeneous resource pool is constructed to reduce the risk of faults caused by a single system. The hot standby function is deployed to implement seamless user switchover in the event of faults and to reduce the signaling storm risk. Cross-domain AI collaboration is introduced, combined with 4G/5G network load, to enable intelligent service guidance and balanced distribution. An end-to-end three-dimensional monitoring and digital twin system is built to implement cross-layer, cross-domain, and multi-dimensional measurement and perception, allowing for prediction and prevention of faults in advance.
- **Transport network:** ZTE focuses on constructing an end-to-end autonomous network that covers the entire lifecycle. The AI and data platform enable proactive network prevention, automatic optimization, and fault self-healing. Closed-loop management functions, including network simulation, AI O&M, transport capability map, and management of service quality differences, have been successfully applied in Shaanxi, Guangdong, Hebei, and Yunnan provinces of China.
- **Wireless network:** To address network risks such as illegal access and stealth attacks, ZTE has integrated intrinsic security features into the design of 5G base stations, including host intrusion detection, virus detection, asset data collection, security event analysis and handling, and security situation presentation, enabling real-time monitoring of the operating status of 5G base stations, rapid identification of abnormal behaviors and security events, and timely implementation of measures to effectively reduce the risks of external security attacks.

Opening a New Era of Digital Intelligence

With the rapid development of generative AI and large models, promoting network service intelligence has become an industry consensus. However, implementing general large models in the communications industry faces many difficulties. In response, ZTE has launched the "Digital Nebula" solution and platform, which enables higher-order intelligent applications in communication network service fields such as intelligent Q&A, intelligent network assistant, intelligent preventive maintenance, precise customer service, and network insight.

The in-depth application of AI and other innovative technologies across various industries will serve as the primary drive for the next wave of digital transformation. In the future, ZTE will continue to work with partners to establish an efficient and intelligent ecological operation mode, build long-term digital and intelligent service capabilities, facilitate the digital and intelligent transformation across various industries, and promote the development of the digital economy. ZTE TECHNOLOGIES

AI and Large-Model Technologies Enable Digital and Intelligent Transformation of Communication Networks

Yang Yahan

Director of ZTE Global Sevice Digitalization Project

Zhou Li

Supervisor of ZTE Delivery Project Management

With the continuous progress of science and technology, especially the rapid development of artificial intelligence (AI) and large-model technologies, the telecom industry is witnessing unprecedented changes. From intelligent network deployment to automated O&M management, AI and large-model applications have deeply penetrated all aspects of communication networks. Communication networks are advancing towards digitalization, intelligence, and automation to better serve the digital and intelligent transformation of the economy and society.

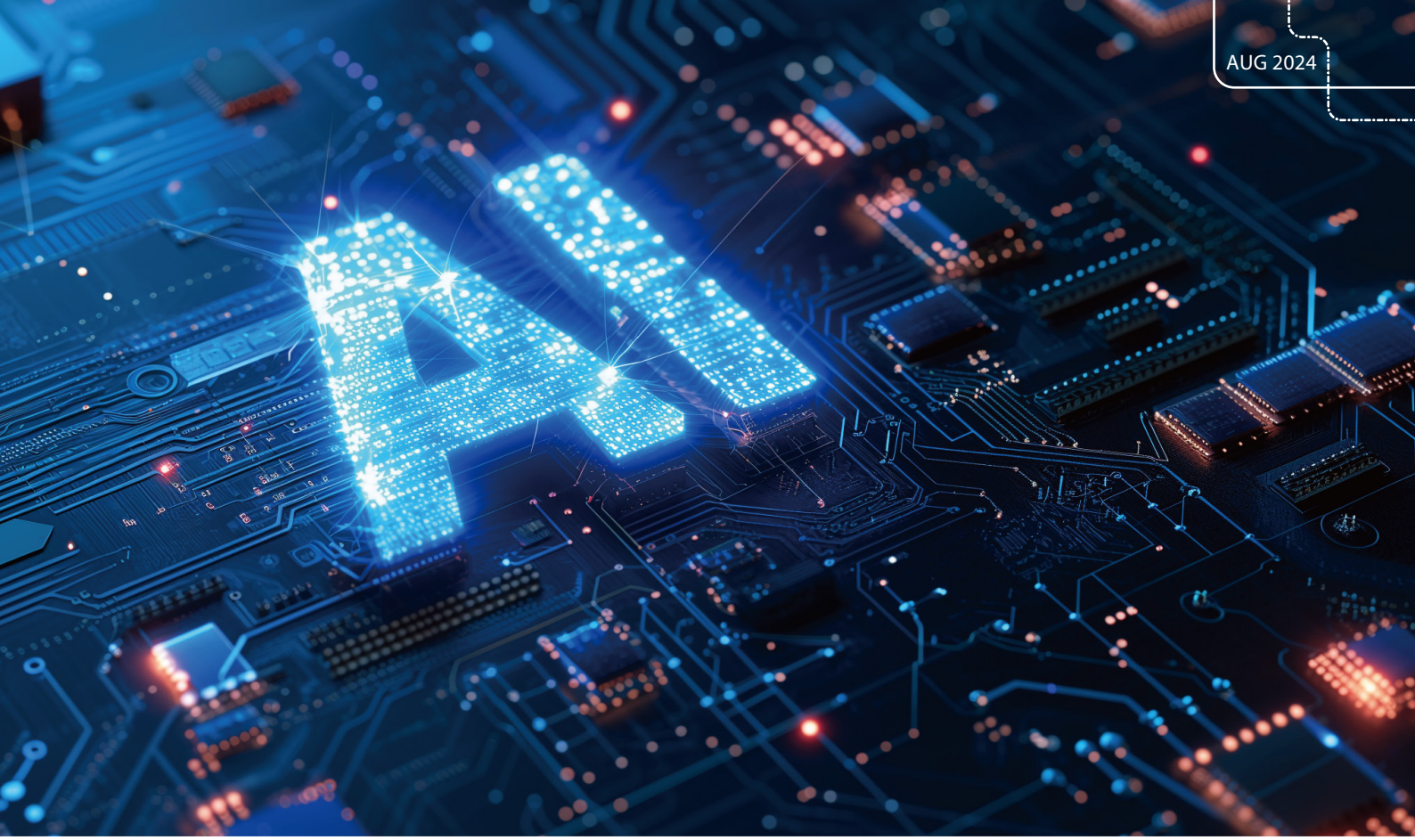
Why Communication Networks Need AI

Mobile communication networks have become increasingly complex in recent years, the communication service ecosystem has diversified, and communication network infrastructure and service systems need to address more complicated scenarios. In these contexts, the traditional manual rule pre-definition and execution processing and management capabilities fail to meet the requirements. Achieving more efficient and intelligent solutions through AI technology has become an inevitable trend in communication network development.

On one hand, AI can rapidly identify and resolve issues by analyzing vast amounts of network data, offering round-the-clock monitoring and management for communication networks through automated monitoring, prediction, and maintenance capabilities. This improves network O&M efficiency and greatly reduces the utilization of human resources. On the other hand, AI can extract useful information from massive data through deep learning and data mining technologies to provide decision-making support and enhance network deployment efficiency. The AI technology can automate and intelligently optimize the utilization of network resources, thereby enhancing network performance, improving user experience, and accelerating the innovation and development of communication networks.

Combination of Communication Networks and AI

From the early stages of automated optimization management, reducing manual intervention, to the



introduction of machine learning for self-configuration, self-optimization, and self-healing, and further integration of AI and large-model technologies as the intelligent brain, communication networks are evolving towards artificial intelligence. The application of AI and large-model technologies in communication networks is primarily evident in the following aspects.

Network Planning and Deployment

Through the analysis of extensive network data, user requirements, and service scenarios, AI and large-model technologies automatically generate optimization solutions to help operators and network equipment manufacturers achieve more intelligent and flexible network planning and design. The data analysis and prediction model can be used to evaluate network requirements more accurately, thereby enhancing the accuracy and efficiency of network coverage and capacity configuration.

The AI image and video recognition technology based on deep learning can automatically detect and recognize network devices. For example, on the network construction site, mobile cameras or dedicated apps are used to capture images of equipment installation and the completeness of safety facilities for construction personnel. The AI technology identifies equipment types and locations, providing real-time feedback on audit conclusions. The accuracy of this technology's identification surpasses manual judgment by a significant margin. This technology not only reduces manual labor but also ensures the correct deployment and connection of devices, thereby guaranteeing high-quality network construction.

Automatic Network O&M

In automatic network O&M, AI applications can monitor and predict real-time network performance while also offering efficient fault diagnosis and repair solutions. This significantly reduces the need for manual intervention and improves network stability and reliability.

Firstly, AI can identify the operation modes of network devices and links through extensive data analysis and learning. It deeply learns historical data to establish a deep understanding of network behaviors. The process is not only limited to the understanding of network indicators, but also includes the recognition of complex relationships

between network devices and the ability to predict potential faults.

Secondly, when a fault occurs in the network, the AI system can respond immediately. For example, if the performance of a network device suddenly degrades or a link is interrupted, the AI system can quickly capture these abnormal signals and locate possible faulty sources based on previous learning and mode identification. This automated fault detection and positioning process can greatly shorten troubleshooting time, ensuring swift resolution of network faults.

Finally, AI can also provide real-time fault diagnosis and repair suggestions. By analyzing the network topology and operational status, the AI system can generate targeted troubleshooting solutions and automatically execute repair operations if necessary. For instance, when the load of a network node is excessively high, the AI system can automatically adjust the load balance to relieve the network pressure and ensure the stable operation of the entire network.

Intelligent Network Resource Management

AI can be used to analyze real-time network traffic, load, and resource usage, automatically adjusting the allocation and scheduling of network resources. This enables dynamic network optimization and maximizes resource utilization.

- **Intelligent bandwidth management:** Some cloud service providers and network operators in China have started using AI to implement intelligent bandwidth management. The AI technology analyzes and manages network traffic, automatically adjusting bandwidth allocation as required to ensure the stability and performance of network services.
- **Capacity planning and prediction:** Some operators and Internet companies have adopted the AI technology in network capacity planning and prediction. They utilize big data and AI technologies to predict user traffic growth trends, optimize network resource configurations and investment plans, prevent network congestion, and improve user experience.

Intelligent Network Security Protection

As network security threats grow in number and complexity, AI and large-model technologies play an increasingly important role in network security protection.

In threat identification and prediction, AI technologies use deep learning algorithms to identify and predict threats within networks. They monitor real-time network traffic, analyze detected abnormal behaviors, and raise alarms to prevent

> AI and large-model technologies are becoming key driving forces in the transformation of global communication network construction and operations towards digitalization and intelligence.

network attacks and data leakage.

In trend analysis and prediction, AI technology can use machine learning to analyze global network attack data, predict future threats based on historical attack models, and provide real-time trend analysis and prediction reports. By analyzing historical data, AI can identify the patterns, behaviors, and targets of attackers, offering corresponding early warning and defense suggestions to ensure network operational security.

Looking into Future

In the future, AI and large-model technologies will accelerate the digital and intelligent transformation of communication networks, enabling more intelligent network management and optimization capabilities, precise capacity planning and prediction, robust fault diagnosis and self-healing capabilities, and advanced network security defense. Through continuous technological innovation, ZTE will facilitate the evolution of communication networks towards autonomous and highly intelligent systems.

- **Autonomous network:** Autonomous network will emerge as an important direction for the future digital and intelligent development of communication networks. At present, communication networks are in the L3 (limited self-intelligence) stage. Utilizing AI and large-model technologies to achieve the ultimate goals of high-level self-intelligence networks and fully autonomous networks is the key for the development of communication networks. However, achieving these goals necessitates ongoing exploration.
- **Application of deep learning and enhanced learning:** In the future, communication networks will increasingly utilize more cutting-edge technologies such as deep learning and enhanced learning for in-depth analysis and understanding of network data and service scenarios. Through the training of large models, ZTE possesses certain awareness and cognition capabilities, enabling flexible solutions to a range of problems and facilitating the improvement of network intelligence and decision-making capabilities.
- **Cross-domain collaboration and ecosystem development:** Operators, equipment manufacturers, cloud service providers, and application developers will strengthen cross-domain collaboration and ecosystem development. They will jointly promote the widespread application and innovative development of AI and large-model technologies in communication network construction and operations through collaborative innovation.

To sum up, AI and large-model technologies are becoming key driving forces in the transformation of global communication network construction and operations towards digitalization and intelligence. With continuous innovation and development in AI and large-model technologies, as well as deepening cross-domain cooperation, technological breakthroughs, and ecosystem development, a truly intelligent and inclusive era is on the horizon. ZTE TECHNOLOGIES

Application of Large Models in Communications Field

Li Ruiming

Chief Engineer of Network Planning & Optimization, ZTE

With the continuous development of 5G networks, the scale of mobile communication networks is expanding, and the co-exsitence of multiple frequencies and multiple systems is making networks more complicated. In the 5G era, alongside the Internet of everything (IoE), various new services are emerging, and there are differentiated ToC and ToB service requirements. A large number of terminals with varying types and capabilities are accessing the networks. Under the traditional network construction and operation mode, the level of network automation is low, the network lifecycle is disconnected at different stages, and the manpower cost associated with "planning, construction, maintenance and optimization" is high. To meet these challenges, automation and intelligence have become critical development directions for future network operation and maintenance.

In 2019, TM Forum introduced the concept of the "autonomous network", which is similar to auto-driving. The autonomous network is divided into six levels (L0–L5) as shown in Fig. 1. At L0-L3, users' requirements, objectives, and constraints can be achieved through policy-driven operations, with requirements transferred through existing interfaces. At L4-L5, the system can adaptively determine its own behaviors through intent-driven interaction, reducing the need for human adaptation. This capability will translate into service flexibility through the introduction of new, customized services without human intervention.

The ultimate goal of autonomous networks is to achieve "full autonomous networks", a consensus that has been widely acknowledged in the communications industry. At present, the industry's autonomous level is generally at L3 (conditional autonomous networks).

In the L3 phase, network operations are automatically performed through pre-defined and pre-tested scripts and policies, greatly improving network O&M efficiency. However, key decision-making points still rely on human experience and require human participation. As the autonomous network evolves into the L4 and L5 phases, the obvious change is to minimize the dependence of network operation on human resources. This requires the network to independently formulate optimal solution and automatically execute tasks. According to the definition in the Autonomous Networks Whitepaper released by TM Forum, an autonomous network is a system managed in accordance with specified objectives or expectations. These objectives or expectations are called intentions, including requirements, objectives, and constraints abstracted in a simplified manner. In brief, the intent is "what" rather than "how". This means the user should tell the system what to do, instead of how to do it, greatly reducing the complexity of work.

Large models, represented by ChatGPT, have unprecedentedly promoted the evolution of intent-driven capabilities of autonomous networks and have led the research and practice within the current industry.

Building Large Model Capabilities

A large model is rooted in machine learning. Machine learning is a branch of artificial intelligence, which refers to the process of providing data to a program to train the program to identify data

features without manual intervention. Machine learning covers multiple algorithms and technologies, such as linear regression, support vector machines, and deep learning based on neural networks. A large model is a machine learning model with a large number of parameters (typically 10 billion or more) and a complex network structure. Such a model is usually pre-trained using large-scale training data and requires a large amount of resources for both training and deployment. Key features of a large model include:

- **Understanding:** Large models well understand human intentions and demonstrate strong instruction-following capabilities.
- **Memorization:** Large models hold multiple rounds of dialogs without forgetting the contents of previous dialogs.
- **Responsiveness:** Large models derive comprehensive user concerns from a large number of user interactions, improve task scenarios, and provide responses that closely match user requirements.
- **Smoothness and logic:** The smoothness of language and the logical rules of large models have surpassed those of most humans.

Since OpenAI released the ChatGPT, large technology enterprises have successively released their own large language models, such as META's LLaMA, Google's PaLM-E, Baidu's ERNIE Bot, and Alibaba's Tongyi Qianwen.

Utilizing the intent recognition capability of large models, the network O&M mode shifts from "How to do" to "What to do", ensuring a fast service experience driven by natural language and free from process interference. For example, in a network fault handling scenario, the function call and API mapping capabilities of large models are used to achieve coordinated orchestration of the structured data models, enabling quick fault detection and alarm association convergence. Based on the output results, large models perform secondary analysis and summarization to generate a fault summary.

In the closed-loop fault handling process comprising solution recommendation, solution implementation, and quality inspection, historical fault cases are retrieved using the retrieval augmented generation (RAG) capability of large models to generate a handling solution. Next, intent identification and transfer are completed through man-machine interaction, and fault self-healing instructions are executed. Then, interactive quality inspection is performed, utilizing natural language to SQL (NL2SQL) for self-querying alarm status. Finally,

Autonomous levels	L0: Manual operation & maintenance	L1: Assisted operation & maintenance	L2: Partial autonomous networks	L3: Conditional autonomous networks	L4: High autonomous networks	L5: Full autonomous networks
Execution	P	P/S	S	S	S	S
Awareness	P	P/S	P/S	S	S	S
Analysis	P	P	P/S	P/S	S	S
Decision	P	P	P	P/S	S	S
Intent/ Experience	P	P	P	P	P/S	S
Applicability	N/A	Select scenarios				All scenarios

P—people (manual) S—system (autonomous)

Fig. 1. Autonomous network levels.

the associated knowledge is extracted through the large model to achieve knowledge recycling and model capability iteration.

Based on the Nebula large telecom model, ZTE has developed a wireless network guarantee assistant that empowers all three phases (beforehand, in-process, and afterward). This assistant quickly generates assurance solutions, implements various assurance tasks including perception evaluation, large-screen monitoring, data statistics and analysis, and cross-domain delimitation and positioning, generates review and summary reports, and recycles knowledge into the database, significantly improving network assurance efficiency in major scenarios and reducing O&M manpower. At the 2023 World Internet Conference Wuzhen Summit, ZTE's guarantee assistant, leveraging the large model, innovated practices for wireless network guarantee scenarios with an end-to-end closed-loop approach and provided generative AI dialog interaction to innovate traditional O&M modes, improving operation efficiency and reducing manpower investment by more than 30%.

Technical Challenges Faced by Large Models and Countermeasures

Compared with traditional methods, large models excel in intent identification capabilities, but still face some challenges in the automatic and intelligent network O&M scenario.

- **Large model hallucinations:** Hallucinations refer to information generated by large models that conflicts with the source or cannot be verified by the available source. These hallucinations may cause large models to generate unexpected output. However, the O&M field has low tolerance for errors, making it crucial to minimize these hallucinations.
- **Weak interpretability:** Large models are recognized as black-box models with complex neural network structures, resulting in generated content that is not easily interpretable. However, the O&M field requires high interpretability of results.
- **Insufficient O&M corpus:** The implementation of large models is hampered by a lack of a large O&M corpus. Especially, there is a shortage of both quality and quantity in private corpus.
- **Combination with existing tools:** It is a challenge to integrate large models with a large number of automatic network O&M tools in the existing network.

To address these challenges, several solutions can be implemented. For example, to mitigate large model hallucinations, we can increase the proportion of explicit knowledge through RAG, utilizing knowledge graphs. To enhance result interpretability, we can employ an "evidence-based" generation strategy. When integrating large models with automatic O&M tools, agents can be used to facilitate these processes.

As a significant technology in the field of artificial intelligence, large models are driving digital transformation across various industries. In the communications field, the natural language processing and intent recognition capabilities provided by large models offer more intelligent and convenient solutions for intent-based autonomous network construction. With the development of technologies and the expansion of application scenarios, large models will play an increasingly important role in the communications field. ZTE TECHNOLOGIES

Innovative Application of AI Technology in Communication Network Planning

Shao Peng

Expert Engineer of Network Planning & Optimization, ZTE

Yan Lin

Expert Engineer of 5G RAN Products, ZTE

With the rapid development of information and communication technologies, artificial intelligence (AI) is gradually penetrating various industries, demonstrating great potential and value in network construction and O&M fields. The application of AI technology in communication networks can significantly improve network intelligence and effectively address many challenges in network construction and O&M. ZTE actively explores innovative practices in intelligent technologies such as AI and large models in the communications field. In network planning, ZTE has introduced an innovative solution, the AI robot for neighbor cell planning. This solution, as a crucial manifestation of AI technology in network planning, addresses the weaknesses of traditional neighbor cell planning applications, providing a brand-new solution for network neighbor cell planning and facilitating efficient network deployment.

What Is AI Robot Needed for Neighbor Cell Planning

In the 2G and 3G eras, neighbor cell planning often relied on manual and tool-based methods. In the 4G era, the introduction of automatic neighbor relation (ANR) technology gradually automated this process. However, the ANR technology requires cell sites to operate within the network for a period before new neighbor cells can be added, and it does not support planning before the cell site accesses the network. As a result, the timeliness of neighbor cell planning is affected.

Tool-based planning involves adding multiple nearby cells to the neighbor cell list in the forward/backward coverage direction of the cell, determined by factors such as the cell's longitude, latitude, azimuth, distance, and angle. However, existing neighbor cell planning methods based on distance and angle have problems such as missing neighbor cell configurations, low accuracy in neighbor cell planning, and an excessive number of neighbor cells, which fail to guarantee the accuracy of neighbor cell planning. Therefore, there is a need for a more timely and accurate neighbor cell planning application to adapt to the increasingly complex communication network environment.

The AI robot for neighbor cell planning innovatively incorporates factors such as the number of neighbor cell layers, distance multiplier, and overlapping coverage area. It can use machine learning to analyze massive neighbor cell data in the existing network, self-learn from the experience of adding neighbor cells in the network, and self-predict new neighbor cells before accessing the network. This implementation enables intelligent and automatic neighbor cell planning, enhancing the timeliness and accuracy of the process.

Self-Analyzing, Self-Learning, Self-Planning Neighbor Cell Planning Solution

As a typical application of AI technology in communication network deployment, ZTE's AI robot for neighbor cell planning comprises several

modules, including basic data extraction, data processing, feature calculation, model training, neighbor cell planning for new sites, and neighbor cell script making. This solution can effectively address the pain points and challenges encountered in each procedure of neighbor cell planning.

Introducing Hierarchical Structure by Using Geometric Knowledge

The AI robot for neighbor cell planning innovatively introduces the concept of Delaunay triangle network. Delaunay Triangles, a common geometric concept, are used to divide a given set of points into disjoint triangles, ensuring no other points lie within the outer circle of each triangle. Utilizing the engineering parameters of cell sites, the robot treats each cell site as a vertex and uses latitude and longitude data to build a Delaunay triangle network. By traversing the adjacent relationships between triangle vertices, the robot can establish the hierarchical relationships between cell sites.

This method, based on the Delaunay triangle, considers the locations of cell sites as a point set. It avoids the creation of simple one-to-one neighbor cell relationships and instead establishes relationships through a hierarchical structure. This approach makes neighbor cell planning more effective, reduces radio interference and handover times, and thus enhances overall network performance and user experience.

Learning Rules of Adding Neighbor Cells to Existing Network Through Massive Data Analysis

AI technologies have the capability of processing vast amounts of data, including both structured data (such as tabular data in the database) and unstructured data (such as text, images, audio, and video). Compared with traditional methods, AI technologies, especially deep learning, can efficiently handle large datasets through parallel computing and distributed processing. They can discern patterns within these datasets and provide decision-making support more efficiently.

The AI robot for neighbor cell planning uses the deep learning technology to extract valuable information from vast amounts of existing network data, including engineering parameters and neighbor cell configuration data. It summarizes the rules of neighbor cell addition, which are usually difficult to discover or understand using traditional statistical methods. This enables the robot to automatically learn and adapt to new data and environment changes. Once the model is established, it can continuously learn new neighbor cell configurations within the existing network and optimize them, thereby steadily enhancing performance and accuracy. This learning mechanism also allows the robot to be applied to different operator networks and to train AI models applicable

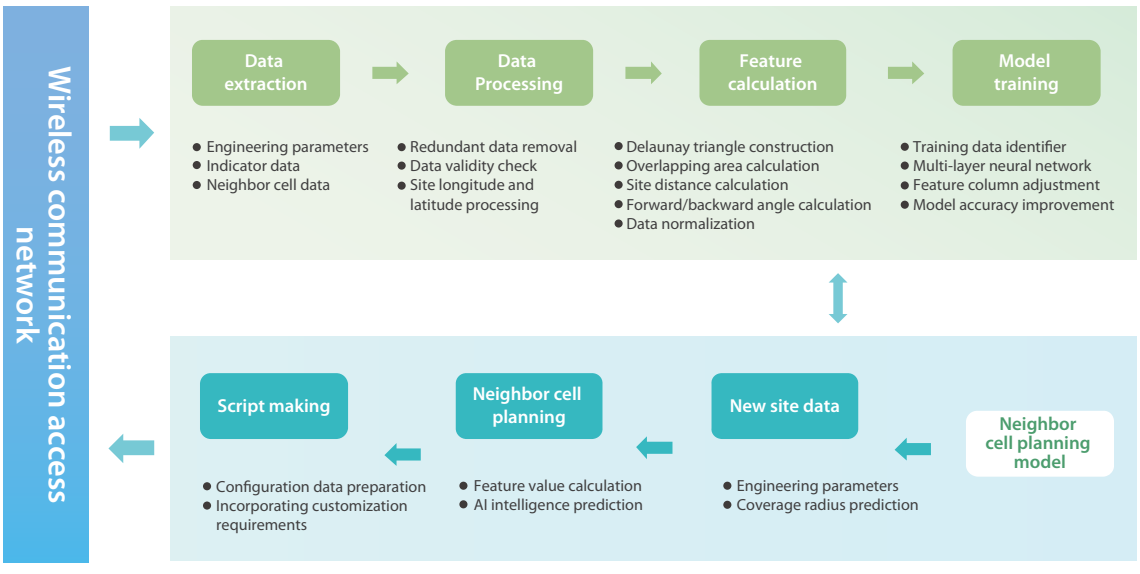


Fig. 1. AI robot modules for neighbor cell planning.

to various operator networks.

Planning New Neighbor Cell Sites Through Intelligent AI Prediction

During the learning process, the AI model will attempt to generalize existing knowledge, meaning it applies the rules learned from training data to new, unprocessed data. This capability allows the model to make reasonable predictions in novel situations, rather than simply memorizing the training data. The neural network is a common model in deep learning, employing a network composed of multiple layers of neurons to model complex data relationships.

Using the multi-layer neural network model, the AI robot for neighbor cell planning reads data from both newly constructed sites and existing networks to make intelligent decisions in complex environments. Once the model completes training and is deployed in the actual network, the robot can predict the best neighbor cell planning solution based on the geographical location and surrounding environment of the new site, aiming to optimize network coverage and performance. This intelligent neighbor cell planning method, based on the deep learning model, significantly improves the efficiency and accuracy of network neighbor cell planning, providing robust technical support and optimization solutions for telecom operators.

Application of AI Robot for Neighbor Cell Planning

With its strong adaptability and accurate planning capability, the AI robot for neighbor cell planning has been widely applied in many overseas projects, providing robust support for intelligent network deployment.

In a large-scale dual-network convergence project in a Southeast Asian country, various network scenarios and site statuses are involved, making neighbor cell planning complex and demanding high computer performance. The AI robot for neighbor cell planning is used to learn the neighbor cell data of the project within the existing network. Leveraging the frequency planning and special rules for neighbor cell addition in the existing network, the time required

for neighbor cell planning and configuration is reduced from 600 minutes each time to just ten minutes. This significant reduction greatly improves network convergence efficiency for the operator.

In an overseas large-scale swap project, neighbor cell planning and optimization scenarios are highly complex, involving intra-frequency, intra-layer, inter-frequency, and inter-system neighbor cell planning and optimization. Traditionally, it takes two weeks to optimize neighbor cells across the entire network each time. However, the AI robot for neighbor cell planning can automatically output key information, such as neighbor cell pair relationships, number of neighbor cell layers, and forward/backward relative azimuth, within a few minutes. The planning can be completed in just a few minutes, and network-wide neighbor cell optimization can be completed within one day, greatly improving the efficiency of network swap and O&M.

Looking ahead, the AI robot for neighbor cell planning will continue to evolve and grow, comprehensively enhancing the intelligence of network neighbor cell planning. This advancement aims to deliver more intelligent and high-quality services and experiences, driving the network towards an intelligent future. ZTE TECHNOLOGIES

AI-Powered Video Perception Solution Enhances O&M Efficiency

Zhang Enyi

Expert Engineer of ZTE Video Technology Support

With the rapid development of broadband and video coding technologies, big video IPTV & OTT services have become fundamental broadband offerings for telecom operators. As these services experience exponential growth in commercial use, the O&M complexity of service systems also escalates significantly. Traditional O&M mode focuses on manual analysis of service statistics reports and alarms, which is inefficient and time-consuming. However, it fails to adapt to the evolving requirements of big video services, such as monitoring performance indicators, node failures, and fault delimiting and locating. Therefore, a more intelligent O&M mode is imperative to improve operational efficiency and ensure the stability, reliability, and quality of big video systems.

AI technologies are introduced into the O&M of big video services, greatly improving their automatic and intelligent capabilities and providing the optimal technical solution for addressing current O&M challenges. On one hand, AI technologies enable the deep integration and unified presentation of various performance and network indicators. On the other hand, AI algorithms can be used to analyze massive and multi-dimensional monitoring data, enhancing the ability for fast fault detection and location.

Responding to the imperative need for enhancing the user experience of video services, ZTE integrates AI technologies to develop the BigVideo ezMon intelligent O&M platform (Fig. 1). It introduces a content distribution network (CDN)-based indicator system, including performance indicators, channel quality monitoring, intelligent scheduling, and log aggregation analysis, to meet the requirements for intelligent O&M in big video networks. Leveraging a big data platform and industry-leading Prometheus and Etcd architectures, ZTE's BigVideo ezMon platform boasts high-performance computing and storage capabilities, professional data processing and analysis capabilities, accurate service quality perception, real-time monitoring and alarm analysis, log aggregation analysis, end-to-end network diagnosis, and intelligent and rapid fault location.

Performance Monitoring for Visualized Indicators

The BigVideo ezMon intelligent O&M platform can collect and monitor KPIs within seconds, offering visualized indicators for various services across multiple dimensions, encompassing the entire network, areas, nodes, and devices (Fig. 2). Users can quickly customize large-screen monitoring indicators and display modes (such as trend charts, histograms, and tables) as required. Through the correlation analysis of alarms and performance indicators, the platform can promptly identify the location of system alarms and their impact on services and users.

- **Service indicators:** In accordance with service monitoring templates across dimensions of the entire network, areas, nodes, and devices, the platform provides various performance indicators such as total number of concurrent users, service success rate, data backhaul success rate, bandwidth usage, delay of first service packet,

request hit rate, and download rate.

- **Hardware indicators:** The platform provides unified hardware monitoring indicators such as CPU load, memory usage, storage usage, NIC IO rate, hard disk IO rate, and NIC packet loss rate.
- **Network indicators:** The platform provides switch monitoring indicators such as interface traffic and packet loss.
- **Health overview:** Through the correlation analysis of alarms and performance indicators, the platform intuitively presents the health overview of various service systems according to dimensions such as the entire network, areas, and nodes.

Channel Monitoring for QoS Assurance

The BigVideo ezMon intelligent O&M platform provides OTT channel monitoring capabilities to obtain real-time QoS data for each monitored channel, quickly identify abnormal services, locate channel faults, and determine their impact scope.

- **Service topology monitoring:** The platform monitors channel service data in real time, aggregates data by node and device, and displays channel QoS on nodes and devices in real time.
- **Channel topology monitoring:** The platform draws the backhauling link based on the actual service of the channel, visually analyzes the service indicators of the nodes and devices on the backhauling link, and quickly identifies abnormal services.
- **Channel quality analysis:** The platform supports the analysis of channel service indicators and impact analysis on areas, nodes, and devices. It visually displays the service change trends within a specific time period, and rapidly locates the impact scope of channel faults.

Intelligent Scheduling for Load Balancing

The BigVideo ezMon intelligent O&M platform provides CDN intelligent scheduling decision-making capabilities. It uses AI algorithms to analyze service data in various scenarios such as intra-area, inter-area, live broadcast hotspots, and VOD hotspots to obtain scheduling policies. These policies are then delivered through automatic interfaces to achieve more accurate and timely intelligent scheduling.

- **Intra-area load scheduling policy:** Based on the load of each node within an area, the platform uses a multi-dimensional dynamic algorithm to provide scheduling suggestions for user groups, node service priorities, and weights, thus implementing the intra-area scheduling policy for multi-node

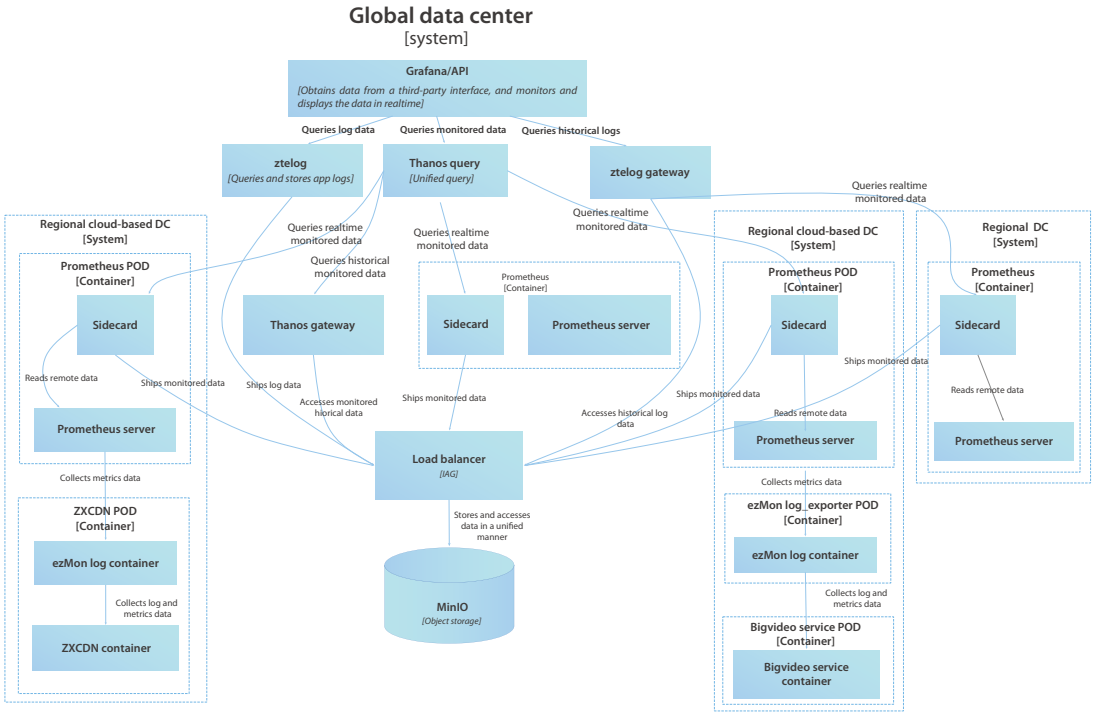


Fig. 1. BigVideo ezMon intelligent O&M platform.

Fig. 2. Network-wide service monitoring.

load balancing.

- **Inter-area load scheduling policy:** Based on the actual load and uplink/downlink load capabilities of each area, the platform uses a multi-dimensional dynamic algorithm to provide scheduling suggestions for user groups, node service priorities, and weights, thereby implementing the inter-area scheduling policy for load balancing.
- **Live hotspot content prediction and scheduling:** The platform analyzes and predicts hotspot live broadcast channels based on the regional dimensions. It then provides distribution policies for these channels, aligning with node service capabilities to facilitate rapid channel creation and rollback scheduling. In addition, it supports configuration of hot channels for special events, ensuring high-priority scheduling for live broadcast hot channels.
- **Hot VOD content analysis and scheduling:** The platform analyzes hot VOD content regionally and provides a distribution policy based on node service capabilities. It can also facilitate rapid VOD creation, duplication, rollback, and cross-area propagation to improve hit rates.

Log Aggregation Analysis for Enhanced Fault Location Efficiency

The BigVideo ezMon intelligent O&M platform remotely collects service and system logs from each device, aggregates and analyzes them, and rapidly retrieves end-to-end service access information. It supports multiple query and filtering conditions, such as IP address and time range, enabling retrieval of process logs from each service module to identify potential exceptions in the entire service process. This enhances the efficiency of fault delimiting and locating.

In a big video project in Turkey, routine O&M involves over 300 servers and more than 30 services, resulting in a heavy workload. Traditional O&M methods could not effectively cover all devices and services within the existing network, leading to passive management and difficulty in identifying operational risks in advance. Addressing these challenges, ZTE deployed the BigVideo ezMon intelligent O&M system. This system enabled real-time large-screen monitoring of network-wide device services through a single portal, facilitating early detection of potential network risks and accurate fault location. This implementation of active O&M significantly reduced the workload of on-site maintenance personnel and improved O&M efficiency by more than 50%. With abundant O&M functions and user-friendly operations, the BigVideo ezMon intelligent O&M system provides simplified O&M experience for front-line personnel and visualizes network quality for customers, winning recognition from both front-line personnel and customers.

As the big video service market continues to expand, operators face challenges such as high user experience expectations and the need for improved O&M efficiency at lower costs. Therefore, intelligent network O&M has attracted much attention. ZTE's BigVideo ezMon intelligent O&M platform solution offers distinct advantages in improving network quality and maintenance efficiency. It is poised to be widely used in future network construction and optimization endeavors, helping customers implement intelligent network O&M. ZTE TECHNOLOGIES

AI-Assisted Security: Creating New Network Defense Strategy with Self-Perception, Self-Diagnosis, and Self-Adjustment

Cao Kunpeng
Senior Expert of ZTE Global Service Security Planning

With the rapid development of network technologies and the proliferation of attack techniques, artificial intelligence (AI) has gradually become crucial in the field of network security to address increasingly complex security threats. In comparison with manual operations mostly based on individual experience and skills, AI offers distinct advantages in network security. Currently, ZTE is actively incorporating AI technologies into its network security solutions and continuously exploring and leveraging their advantages to maximize defense against increasingly complex network attacks and threats in new situations.

After years of dedicated development, AI has played an important role in ZTE's products and network security solutions, enabling features such as intelligent O&M control, automatic data configuration, endogenous security, and pseudo base station detection.

Intelligent O&M Control: Minimizing Human Errors

Network O&M involves numerous complex and high-risk operations, occasionally resulting in security incidents due to incompetence, manual negligence, or malicious intention. To effectively address such incidents caused by human error, ZTE has innovatively developed the iNet system, harnessing digitalization and AI technologies (Fig. 1). This transformation shifts operational status from "offline" to "online" and transitions from "human-reliant control" to "machine-reliant control". By establishing a unified operation platform and

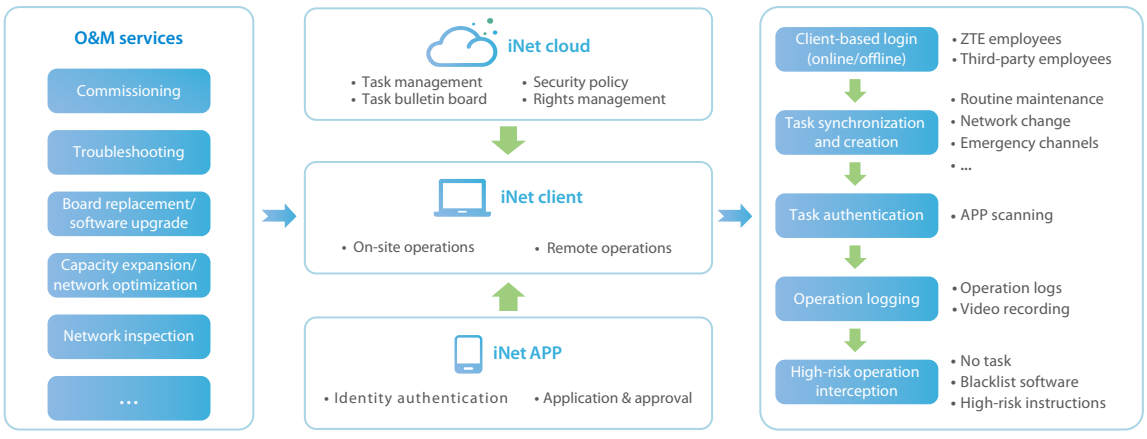


Fig. 1. The iNet system.

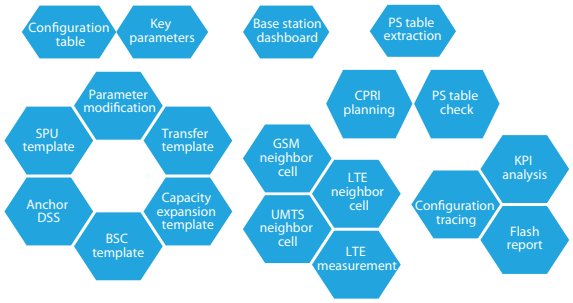


Fig. 2. Automatic base station configuration by robots.

entry, the iNet system not only avoids siloed operations but also enhances network security capabilities, including standardizing operations, preventing unauthorized activities, intercepting high-risk operations, and facilitating auditing, through digitization of the entire operational process before, during, and after operations. The iNet system also uses identity authentication, APP verification codes, permitted operation periods, whitelisted tools, and automatic interception of high-risk instructions to ensure that only the "right person" can perform "right operations" at the "right time" based on the "right task". Additionally, the system supports security traceability through automatic screen recording and operation log collection to record all activities.

Digital Robots: Improving Configuration Efficiency and Reducing Human Errors

To address issues such as low efficiency and accuracy caused by extensive manual parameter settings in complex network construction, ZTE has launched a robot solution aimed at implementing automatic, collaborative, and efficient configuration of advanced wireless base station devices (Fig. 2). The configuration data and key parameters modules of the robot facilitate rapid and accurate extraction of intricate network data, thereby adequately preparing for subsequent network design and planning. Addressing challenges such as time-consuming manual configurations and frequent human errors, the robot provides functionalities like automatic topology planning and template filling for base stations. Additionally, to facilitate base station monitoring, the robot offers a visual monitoring solution incorporating key alarms and indicators. With the widespread deployment of robots, manual operations are greatly reduced, leading to Zero-Wait-Time-and-Zero-Error service experience, and significantly improving the deployment efficiency of communication networks.

Endogenous Security: Prompt Threat Monitoring and Isolation

The overall endogenous security-based attack perception framework of ZTE 5GC product is as shown in Fig. 3. This framework comprises two key components: attack perception center and micro-isolation policy management center. The

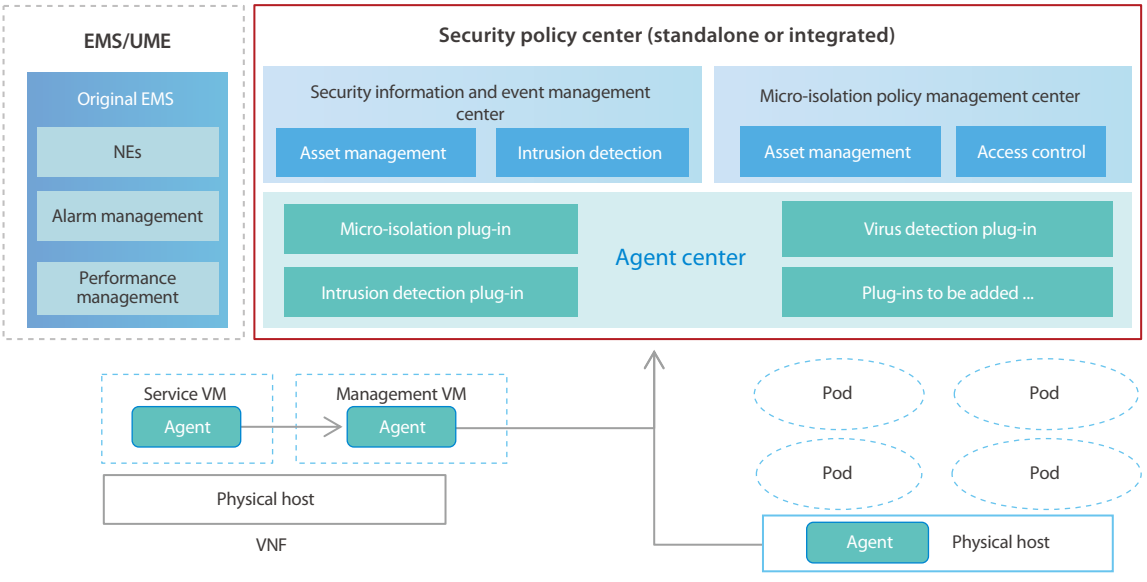


Fig. 3. Endogenous security-based attack perception framework of ZTE 5GC product.

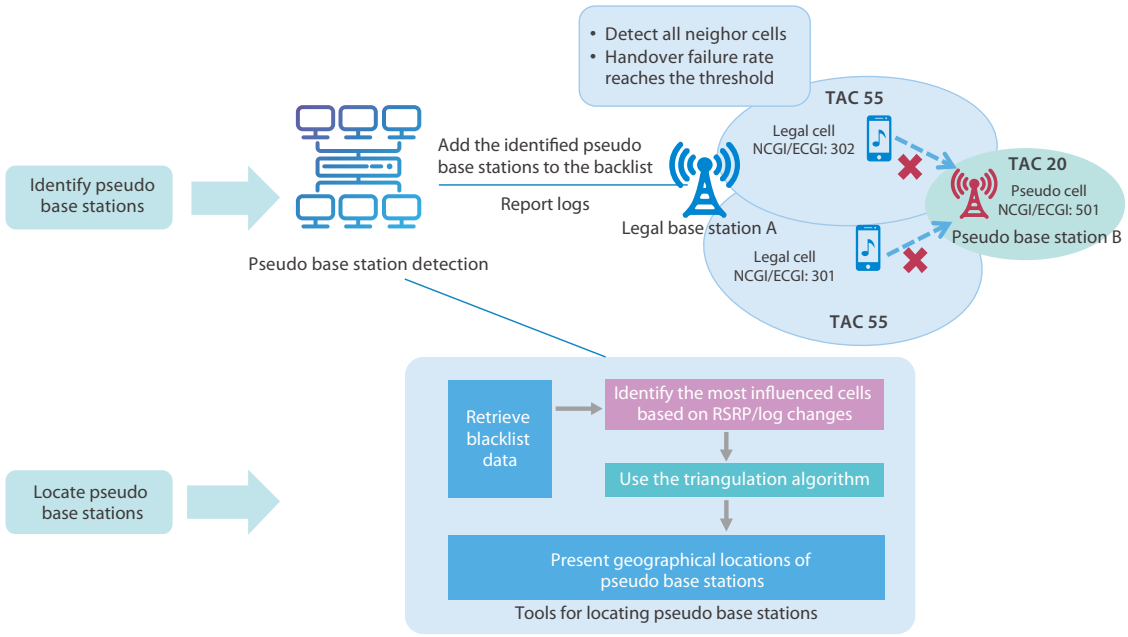


Fig. 4. Pseudo base stations detection scheme.

attack perception center provides the intra-NE attack perception function to promptly detect various abnormal behaviors within the NE. Moreover, employing the attack perception center enables lighter attack perception components, better meeting carrier-class requirements for high performance and high reliability. The micro-isolation policy management center supports the micro-isolation function, marking locations of compromised hosts and suspected ones, persistently conducting real-time monitoring of threat scopes, and subsequently issuing alarms and recording logs. After manual correction, fine-grained security access control is further implemented on the previously identified abnormal ports and connections.

Pseudo Base Station Detection: Ensuring Reliable Service Operations

Pseudo base stations, deployed by malicious attackers, pose prevalent threats in many countries. These devices act as legal mobile base stations to steal subscribers' communication data, track locations, and carry out phishing attacks. Pseudo base stations not only compel mobile subscribers to disconnect from the original public mobile communication network but also disseminate false information to the public, disrupting social order and potentially endangering national security. Leveraging the unique features of pseudo base stations, ZTE has devised a detection scheme to collect and analyze abnormal and associated information, triggered correspondingly (Fig. 4). This scheme enables the determination of pseudo base station locations and relevant details based on the affected real cells/UEs, and presents this information geographically. Consequently, pseudo base stations can be accurately identified and located. Furthermore, the identified pseudo base stations are added to a blacklist to mitigate their impact on services.

As security threats become increasingly complex and the conflict between attacks and defenses intensifies, AI offers innovative ideas and methods to efficiently ensure network security. Its unique ability to automatically learn, monitor, analyze, diagnose, adjust, and adapt maintains strong momentum in its development within the network security field.

In the future, ZTE will attach more importance to network security management, coordinated defense, and personalized security protection as AI is gradually incorporated into the entire network life cycle. Also, ZTE will work with network operators to enhance security assurance capabilities more intelligently. ZTE TECHNOLOGIES

AI Boosts Intelligent Upgrade in Communication Engineering Quality Inspection

Jia Jia

Chief Quality Management Engineer, ZTE

Artificial intelligence (AI) is a scientific technology used to research, develop, and apply techniques for simulating, extending, and expanding human intelligence. The ultimate goal is to enable machines to possess human-like intelligence and continually evolve. With the ongoing evolution of computing capabilities and technologies, AI has been widely applied across various industries. For instance, the visual technology employing AI deep learning can achieve intelligent inspection and review, significantly enhancing inspection efficiency in scenarios such as vehicle and pedestrian detection on streets, facial recognition at railway stations and airports, and industrial quality inspection on production lines.

Challenges

For the delivery quality management of communication network projects, especially wireless base station projects, subcontractors typically take photos on-site with their mobile phones. Subsequently, the project quality team manually reviews these photos, resulting in overall low efficiency. Could AI deep learning serve as an alternative solution to manual auditing?

Yes, it can. However, when compared with the single scenario of facial recognition at railway stations and airports using fixed cameras, developing the intelligent auditing function at actual wireless base station delivery sites faces the following technical challenges:

- The quality of photos is uncontrollable, as images captured are easily affected by lighting conditions, hand tremors, motion blur, and other factors.
- The inconsistency in camera distance and angle affects the size and proportion of the objects.
- There are multiple quality checkpoints in one photo, and their distribution is unbalanced.
- The convergence of models exhibits a black-box nature, significantly impacting the accuracy of AI model recognition.
- Detection of communication equipment requires high-level expertise, while visual algorithms involve some level of abstraction, resulting in significant communication costs for technical and business teams.

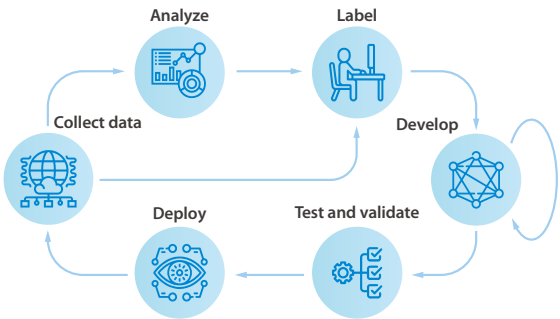


Fig. 1. AI ring R&D process.

Innovative Solution: Eliminating Pain Points

In response to these challenges, ZTE has been continuously innovating, providing a series of innovative solutions to facilitate intelligent quality inspection at communication engineering sites.

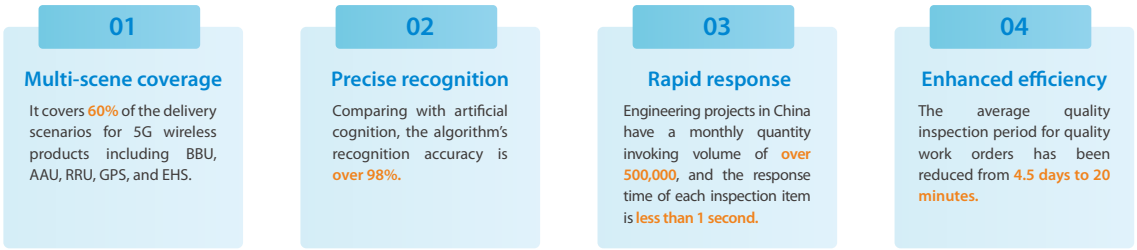


Fig. 2. AI quality inspection effect.

Technically, leveraging deep learning-based target detection and instance segmentation neural network technologies based on multi-scale feature extraction, it can capture rich details of photographed objects under different shooting conditions, addressing the uncontrollability of mobile phone photography. ZTE also utilizes the class activation map (CAM) technology to analyze the discriminative degree of different image categories through CAM heat maps, adjusting model structure and loss functions to address the issue of uneven distribution of multiple inspection points. Drawing on practical experience, ZTE has developed the in-house EasyImage tool and established the AI ring R&D process (Fig. 1) to offer diverse solutions for incubating needs across various dimensions such as recognition accuracy, business relevance, and expertise, thereby significantly enhancing the automation of the R&D process.

- Providing out-of-the-box, sub-second response visual algorithms.
- Offering collaborative online annotation for multiple users, addressing the challenges of ambiguity and metric fuzziness in algorithm R&D through visualization of annotation results and model metrics.
- Leveraging a data lake to establish a closed loop for data and image samples within the service system, providing online analysis and reporting capabilities to aid in service decision-making and algorithm optimization.
- Integrating AI GPU with the iEPMS digital delivery platform to achieve end-to-end online data closed loop, replacing manual, simplistic, and repetitive tasks.

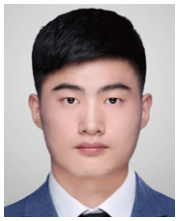
Intelligent Quality Inspection: Enhancing Quality and Efficiency

ZTE has completed the research and development of 17 types of quality inspection algorithms, equipped with capabilities such as multi-scene coverage, precise recognition, rapid response, and enhanced efficiency (Fig. 2). Leveraging the digital delivery platform iEPMS and scenario-based AI algorithms, ZTE binds AI models to inspection scenarios, using AI auditing to replace manual auditing. This enables instant inspection and auditing for quality work orders, greatly shortening the audit cycle. The average inspection period for work orders has been reduced from 4.5 days to 20 minutes, significantly cutting labor costs, minimizing the number of repeated site visits by partner personnel, and enhancing the efficiency of network construction. At the same time, it comprehensively ensures the quality of network construction, facilitating efficient network delivery.

The integration of AI technology into communication engineering scenarios marks a significant step toward automating high-frequency repetitive tasks, thereby replacing manual labor. This not only fosters innovation but also reshapes traditional industries and generates new operational models. Moreover, sharing AI models facilitates collaboration and knowledge exchange with customers, propelling technological development and industry progress.

In the future, ZTE will continue to explore the utilization of AR glasses, cloud rendering, spatial computing, and coordination capabilities to enable more innovative application scenarios such as AI inspection, AR acceptance, and AR training. Leveraging emerging intelligent technologies, ZTE aims to optimize business processes, facilitate the deep integration of digital technology with physical operations, collaborate with partners for intelligent and efficient ecosystem operations, deliver high-quality networks to customers, and achieve greater value realization. ZTE TECHNOLOGIES

Digital Tools for Engineering Survey and Design of Communication Networks



Zhang Bolun

Engineering Design Engineer of ZTE Engineering Delivery Dept.

To adapt to the rapid development of the global digital economy and speed up the construction of communication networks, engineering survey and design services are undergoing a digital and intelligent transformation. Relying on the self-developed digital delivery platform iEPMS, ZTE has launched the electronic technical site survey system (eTSS) to enhance the efficiency of survey and design services. This system provides digital and intelligent solutions for the rapid and high-quality delivery of communication network projects.

As the front-end of network delivery services, the accuracy and completeness of survey directly affect the implementation of design solutions. The traditional survey mode exhibits the following issues:

- **Non-standard survey data entry:** Incomplete survey data and poor photo quality often lead to rework, significantly affecting both project execution efficiency and the implementation of back-end solutions.
- **Long survey report submission period:** In the traditional survey mode, after completion of the survey, the technical site survey report (TSSR) needs to be manually created based on survey information, resulting in low production efficiency and delayed submission.
- **Low efficiency in review management:** In the traditional approach, survey reports are typically submitted offline for review. Communication regarding survey progress often relies on emails or phone calls, which is time-consuming, labor-intensive, and results in high management costs.

ZTE's eTSS system focuses on the above-mentioned pain points, achieving full-process digital and intelligent management.

- **Full process online handling:** The eTSS system achieves full-process online digitization of survey and design services, from creating survey fields to submitting survey technical reports. It effectively addresses the issue of real-time visibility for site survey task status and report submissions, supporting collaborative and visual management of survey and design services for customers, partners, and ZTE.
- **Efficient task management:** The eTSS system supports data synchronization and interaction between apps and the web client. Survey and

Fig. 1. Wide applications of ZTE's eTSS system.

design tasks can be edited online, while task reporting and approval can be viewed and managed in real time. Task-related report data can be output with a single click, and information is displayed in multiple dimensions, improving the efficiency of managing survey and design tasks.

- **Easy operation:** The eTSS system features a simple and intuitive interface with a clear function layout, reducing user cognitive load. The operation environment is simple and convenient. Partners can complete individual sit survey tasks using the mobile APP, thereby lowering the required level of professional expertise for personnel.
- **Multi-account collaboration for efficiency improvement:** The eTSS system provides multi-account collaboration and allows task assistants to be added as needed for site survey tasks. It simultaneously processes data provided by both the responsible person and assisting personnel to improve on-site operational efficiency.
- **Invoking historical data:** The eTSS system supports the invocation of historical data, so that site historical data from previous projects can be fully utilized in subsequent ones. When surveying the same site in later projects, surveyors only need to check for updates in the historical data instead of conducting the survey again. This functionality reduces on-site workload and enhances the efficiency of survey tasks.
- **Automatic report output:** The person in charge of a site survey task can collect data via the mobile app and generate structured TSSR documents with a single click. This function shortens the time required to produce site survey reports, greatly improves TSSR submission efficiency, and facilitates efficient approval and acceptance of survey documents.

At present, ZTE's eTSS system has been successfully applied in over 120 projects worldwide (Fig. 1). By the first quarter of 2024, it had processed nearly 15,000 survey documents. In an Asian country's network delivery project, the cycle from survey completion to report submission was shortened from an average of 8.9 days to about 2.5 days, increasing the efficiency by 72%. In an African project, a partner uses a single mobile phone to complete all survey and report submission tasks without requiring additional manpower for offline survey reports. This significantly enhances partner on-site survey efficiency and effectively supports efficient network delivery.

In the future, ZTE will continue to explore the field of electronic engineering surveys. With advanced AI technologies, ZTE will closely integrate survey and design scenarios with AI models to promote the automation and intelligence of survey and design services. This will enable efficient delivery of communications network projects and foster collaboration with partners to create greater value for customers. ZTE TECHNOLOGIES

AI Drives Intelligent Integration of Engineering Material Management

Wang Yan

Expert Engineer and Data Director of ZTE Global Service Digital Planning

In the rapidly developing digital economy era, the rise of artificial intelligence (AI) has brought unprecedented opportunities to industries. As a global leader in communication and information technology solutions, ZTE closely tracks AI technology trends and actively explores innovative applications in material delivery for communication network engineering.

In communication network delivery projects, equipment assets serve as the core carriers of contracts and orders. Traditional asset management and delivery methods often involve multiple turnover points, scattered storage locations, various product types, unbalanced supply and demand, and mismatched supply and demand plans, all of which significantly affect project delivery efficiency. Therefore, ZTE proposes an innovative solution that intelligently integrates digital and physical aspects. Through digital business transformation, automated process orchestration, AI inference models, and machine learning algorithms, this solution achieves full-cycle management of material delivery. It includes accurate delivery, efficiency improvement, inventory reduction, intelligent prediction, and intelligent decision-making. Furthermore, this solution supports decision analysis and optimal estimation of production and delivery plans before orders are put into production, enabling complete, highly transparent, and fast-response material management services.

Digital Material Delivery Process

For global delivery projects, ZTE has deployed digital capabilities related to material requirements, physical material transfer, and project material delivery flows on the iEPMS digital delivery platform. This deployment aligns with the material delivery process and key user roles in communication engineering projects.

Automatic Orchestration of On-Site Material Requirements

To address the high-frequency and low-efficiency issue in traditional manual material requisition processes, ZTE has deployed two functions on the iEPMS platform: engineering bill of material (EBOM) and electronic delivery order (EDO).

In managing EBOM operations, ZTE implements extended data design and supports customized configuration to accommodate various service scenarios and requirements. The extended data design encompasses different warehouses and storage locations, different packaging modes, material substitution relationships, contracts, and priority distribution rules. These features provide the system with conditions for automated data analysis.

When managing EDO operations, ZTE integrates inventory information, contract order scope, and EBOM list from upstream service links. Using automated orchestration and completeness analysis algorithms, it achieves batch automatic DO generation across multiple sites, and automatically splits bills based on different warehouses, material sources, and sites. Intelligent orchestration and optimization throughout the material requisition, outbound allocation, and site resource allocation

processes enhance efficiency in preparation and picking for front-line engineering and warehouse personnel.

Modular Construction for On-Site Material Transfer and Collection

Combining with the key path of on-site material transfer in communication projects, ZTE has independently developed a series of functions including site receipt (SR), warehouse receipt (WR), delivery return (DR), delivery moving (DM), stocktaking (ST), reconciliation, inventory query, and transaction query. These functions enable material data collection through mobile apps, and integration of packing details, ex-warehousing documents and total design quantities. This allows operators to perform self-signing, self-checking, immediate review and correction, and discrepancy checking at the construction site.

Moreover, upstream contract configurations, delivery batches, and turnover records are meticulously traced to achieve real-time visualization and precise global material turnover tracking. Embracing an intelligent and simplified approach, ZTE ensures meticulous and efficient management of on-site material flows, facilitating automated project delivery, acceptance and on-site risk management and control.

Visualized Material Delivery Cycle and Predicted Application of Machine Learning

During the full cycle of material order delivery, in addition to on-site delivery management, ZTE also manages and deploys material delivery through the supply chain management (SCM) cloud platform to ensure timely and complete delivery, smooth order performance, and prevent inventory redundancy.

ZTE divides the configuration, production scheduling, warehousing, delivery, customs affairs, delivery, and other services related to order delivery into different operational data languages. It uses "material code" as the master data, converting it into other operational languages to establish a data mapping relationship that connects the entire lifecycle of materials and visualizes their complete delivery process (Fig. 1).

To further support the reasonable formulation of project delivery plans and reserve a sufficient handling period for risks, ZTE utilizes the linear regression algorithm to model global historical logistics routes. It continuously trains and optimizes the prediction model to accommodate varying logistics periods across different countries around the world, thereby predicting more accurate arrival time. This helps in formulating reasonable project plans and dealing with material shortage or surplus risks promptly, ensuring the achievement of project delivery objectives and meeting customer requirements.

AI Operations in Decision Analysis for Delivery Planning Scenarios

The uncertainty in production scheduling, freight, and customs clearance periods may lead to inconsistencies between material supply and the engineering delivery plan. Therefore, ZTE employs scenario-based splitting and applies AI to accommodate both long and short delivery periods in the engineering and material supply plans.

- **Short-term and small-cycle scenario:** ZTE predicts

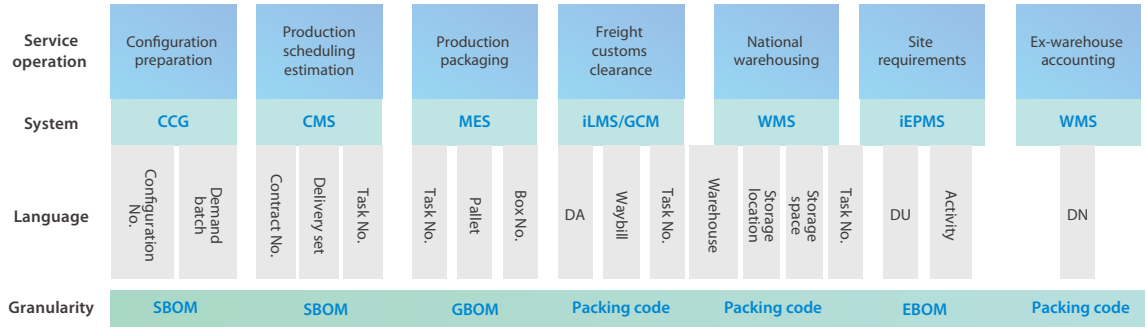


Fig. 1. Visual data diagram throughout complete material delivery cycle.

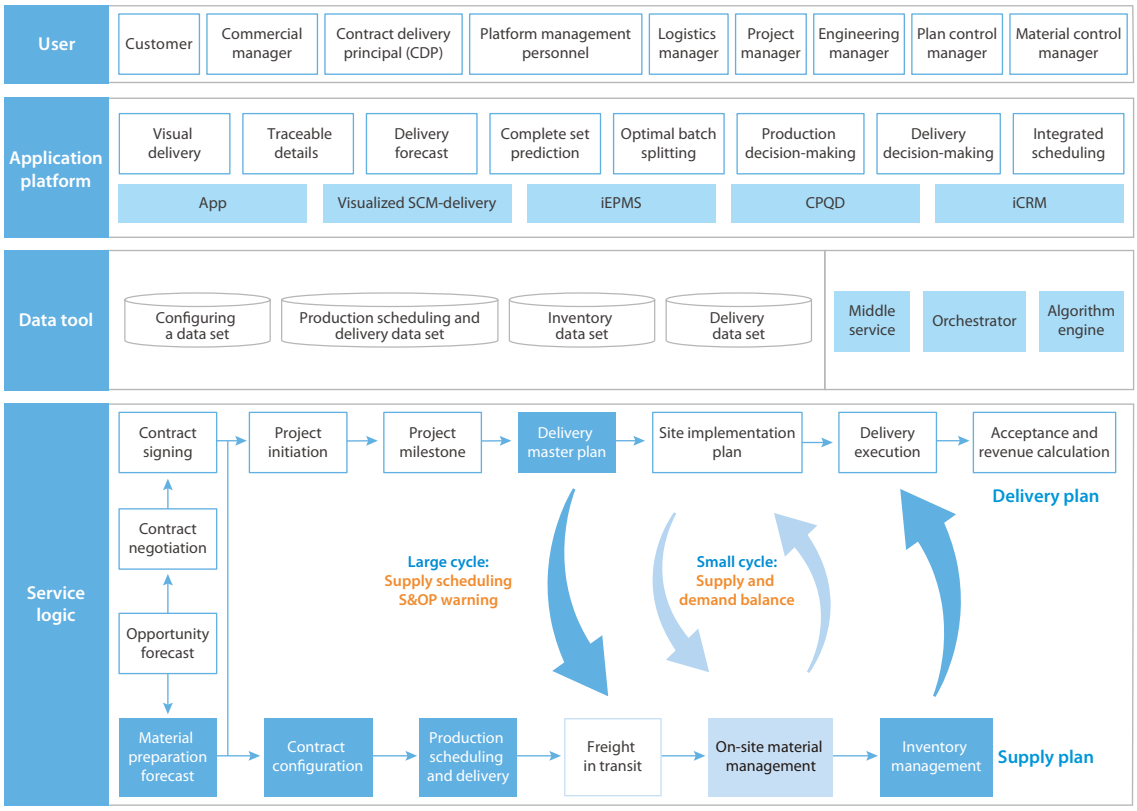


Fig. 2. Intelligent material forecasting for large and small cycle scenarios.

supply and demand balance within one to three months. Based on specified site delivery requirements and estimated arrival time, ZTE intelligently forecasts material needs using a time series model. It uses the decision tree to solve the resource allocation problems across various projects and sites, thereby reducing labor costs associated with on-site material shortage predictions, complete delivery plans, and batch adjustment plan diagnoses.

- **Long-term and large-cycle scenario:** The objective is to deal with material preparation, inventory turnover, and scheduling over a period of three to six months or longer. Building upon the overall supply and demand balance algorithm, ZTE incorporates various uncertainties and potential input variables. It forecasts long-term material requirements using associated material regression prediction and multi-variable time series analysis. Furthermore, it supports decision-making regarding material production, demand, delivery, and optimal plan recommendations based on operational algorithms. ZTE is also exploring the continuous training and optimization of AI models for multi-variable scenarios to enhance the accuracy of prediction and optimal values.

As shown in Fig. 2, through the mutual nesting of large and small cycle scenarios, ZTE can address inventory turnover and material scheduling issues at various time scales throughout the project lifecycle, thereby improving the efficiency and resilience of material delivery.

At present, ZTE's intelligent integration solution for engineering materials has been applied in more than 45,000 projects worldwide, serving over 2 million users. It manages a cumulative total of 800 million on-site delivered materials, significantly enhancing on-site material delivery efficiency while reducing consumption and costs.

In the future, ZTE will continue to explore innovative applications of the AI technology in the communications field, accurately identify service scenarios, improve the delivery efficiency throughout the project lifecycle, inject new vitality into communications network construction, and help customers continuously enhance network value. ZTE TECHNOLOGIES

AI-Empowered ZTE SBA Delivery Modes Facilitate High-Quality AIS Network Deployment in Thailand

Li Yong

Chief Project Quality Engineer, ZTE

Thailand is currently experiencing rapid digital transformation, with a growing demand for the expansion and upgrade of communications infrastructure. AIS, the largest mobile operator in Thailand, has launched a large-scale network replacement and capacity expansion project to maintain its market leadership and meet the surge demand for mobile data. This launch imposes higher requirements for the speed, coverage depth, and quality of network construction. However, as the project progresses, a series of challenges have emerged, including shortages in human resources, the need to enhance team skills, rapid deployment, wide distribution of sites, high quality standards, and strict requirements for problem handling speed.

As a world-leading provider of integrated communications and information technology solutions, ZTE, with deep industry experience and strong technical strength, has innovatively introduced the Simplified execution+Best practices+AI empowerment (SBA) delivery model for quality management and control. This model integrates risk database identification, promotes best practices, applies AI computing tools, and has successfully assisted Thailand AIS in completing its major network replacement and expansion project. This achievement fulfills the goal of secure, rapid, in-depth, and high-quality network construction.

Simplified Execution with Risk Map and SOP

Different projects have different risks, necessitating end-to-end identification and response to ensure high-quality project delivery. In the early stage of the AIS project delivery, the project team worked out the end-to-end process in accordance with the ZTE global delivery risk map and hundreds of local implementation risk points. They identified project-based control points and implemented key resource investments to guarantee and prevent control for high-risk items (Fig. 1). These included team training, service scenario identification, and solution verification, supported by vivid training documents with pictures and videos, configuration collection-translation-verification tools, mirroring environment test, pre-debugging, and service verification tools. This refinement aimed to optimize SOP and management regulations,

striving to achieve targeted control and meet requirements effectively.

Best Practices for Quality Prevention to Improve Efficiency and Quality

The skill of project team members, especially the capability of the partner team, is crucial for successful project delivery. The Thailand AIS project adopts the best team construction mode and best practice process through workshops to change project delivery from disorderly to orderly parallel operations and cross-check process controls. This improvement enhances the efficiency and effectiveness of project delivery.

To improve team capabilities, a combination of best practices, vivid pictures and videos, and mini workshop training has been employed. This shift from one-way training to interactive bidirectional workshops encourages new teams to reflect on their performance, thereby improving training effectiveness. As a result, the time required for a new team to be fully independent and qualified for construction has been reduced from 20 days to 7 days.

To effectively improve the speed and quality of configuration collection during project delivery, and to effectively utilize various types of data, ZTE has self-developed a project-based configuration collection and translation tool. This tool automatically extracts and translates network service configurations, ensuring the accuracy of configuration scripts and operational security. At the same time, a service verification tool can automatically verify 1:1 service duplication before and after cutover (Fig. 2). The application of these tools has increased the efficiency of the whole service cutover chain by 660% compared with the manual mode. In a single day, the number of cutover sites for an IP project exceeds the industry relocation speed, thereby reducing service interruption time and significantly improving end

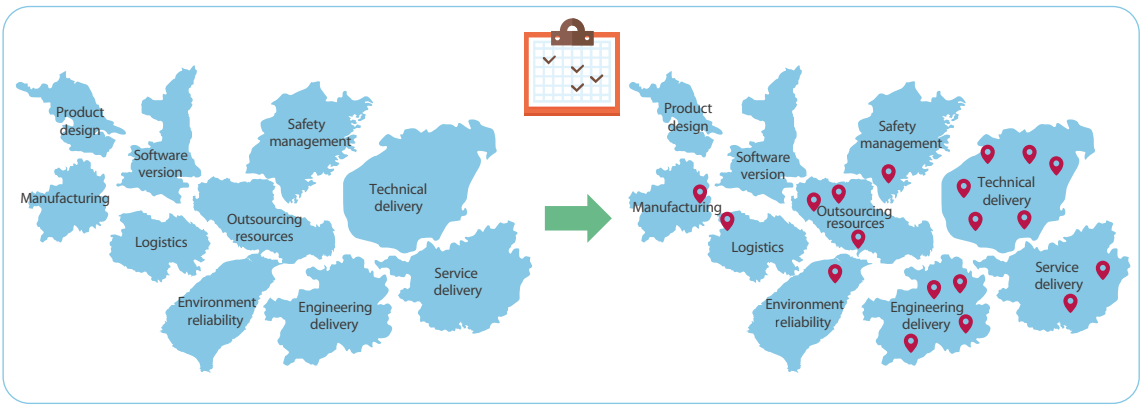


Fig. 1. Targeted project-based control points.

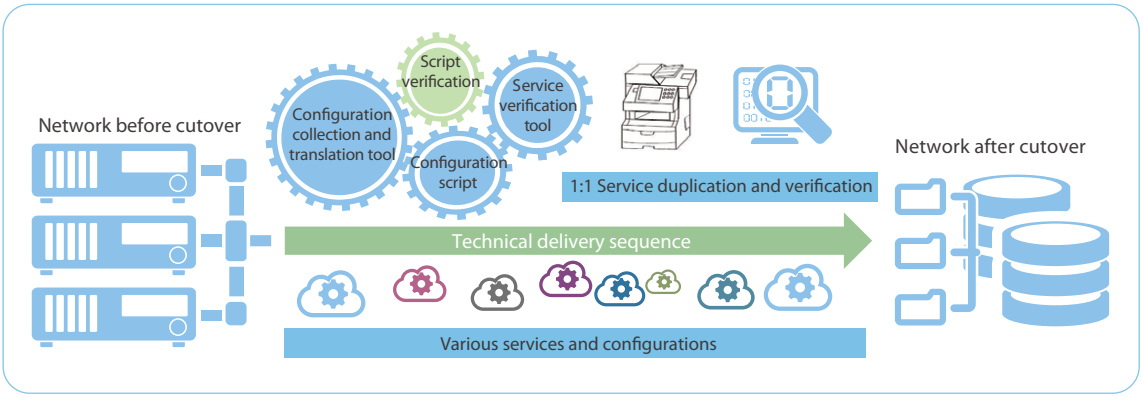


Fig. 2. Series of tools for quality guarantee and efficiency improvement.

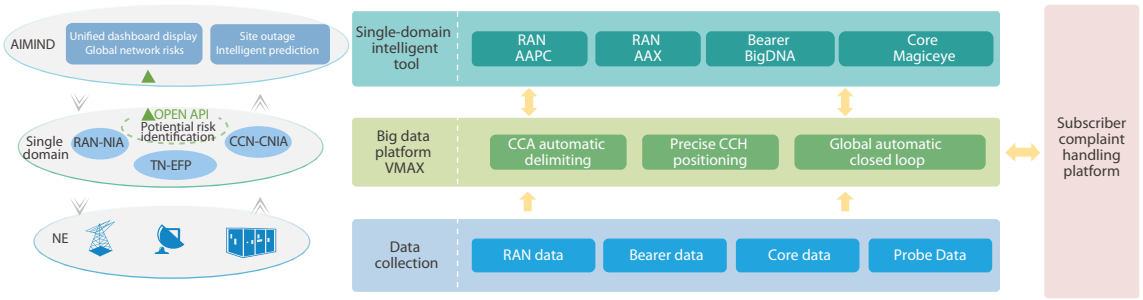


Fig. 3. Self-intelligent network cross-domain coordination and optimization.

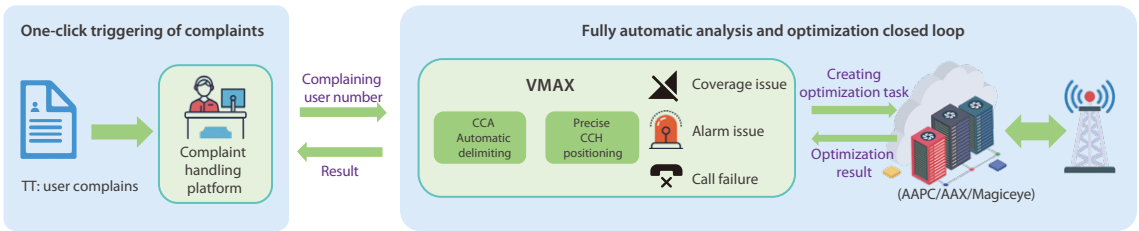


Fig. 4. Automatic delimiting and handling of subscriber complaints.

user perception.

AI Empowerment for Safe Production and Intelligent Operations

During peak periods of the project, more than 200 delivery teams work daily across various sites for high-altitude operations. The innovative ZTE automatic inspection system implements automatic audit management, such as real-time auditing via AI computing cloud, rectification reminders, and approvals, thereby fully ensuring efficient and safe project execution.

The intelligent OSS tool AIMIND integrates data, including historical alarms, performance, and resources from wireless, core, and bearer networks. This enables synchronization of key hardware information across various domains such as links, optical modules, voltage, and temperature. Through AI self-learning, AIMIND establishes a typical fault prediction model to predict sites out of service and faults in wireless base station NEs so as to minimize their impact. In addition, through interconnection with intelligent tools in these domains, it enables end-to-end analysis and fast closed-loop response to complaints, effectively improving network quality (Fig. 3).

In the field of network operation and maintenance, leveraging big data and computing power, the VMAX tool connects network element data related to wireless, bearer, and core networks. It realizes one-click delimitation of problem areas such as terminals, wireless network, transmission network, core network and SP. Furthermore, it implements end-to-end automatic optimization for coverage issues and faults, incorporating self-intelligence functions within individual domains (Fig. 4).

With the innovative SBA integrated quality control management mode and the application of multiple advanced tools, ZTE has successfully achieved important milestones one month ahead of schedule and won the top network quality performance in the third-party network assessment. In addition, the world's latest slicing solution deployed by ZTE lays a solid foundation for the further evolution of 5G technologies in AIS.

Looking ahead, ZTE will continue to uphold innovative ideas and keep improving its technical strength and service level. Amidst fierce competition in the global communications market, ZTE aims to deliver higher quality and more efficient communications solutions and services to customers, contributing to ongoing prosperity of the telecommunications industry in Thailand. ZTE TECHNOLOGIES

ZTE

To enable connectivity and trust everywhere