# GTI 5G-A Wireless Network Intelligence Evaluation System White Paper

GTI

http://gtigroup.org/

# *GTI 5G-A Wireless Network Intelligence Evaluation System White Paper*

| | |
|---|---|
| **Version:** | v_1.0.0 |
| **Deliverable Type** | ☐ **Procedural Document**<br>☑ **Working Document** |
| **Confidential Level** | ☑ **Open to GTI Operator Members**<br>☑ **Open to GTI Partners**<br>☐ **Open to Public** |
| **Program** | 5G Technology and Product |
| **Project** | Technology Evolution |
| **Task** | 5G-A Technology |
| **Source members** | China Mobile, Huawei, ZTE and VIAVI Solutions |
| **Support members** | Ericsson |
| **Editor** | Jiachen Zhang(CMCC), Lei Cao (CMCC), Tianming Jiang(CMCC), Jing Gao(CMCC), Zhe Li(Huawei), Xiongqinshui Gan(Huawei), Pan Li(ZTE), Donghong Ouyang(ZTE), Jianguo Li(ZTE), Jun Yang(Viavi), Yaning Ma(Viavi) |
| **Last Edit Date** | 2024-11-18 |
| **Approval Date** | |

# Document History

| Date | Meeting # | Version # | Revision Contents |
|---|---|---|---|
| | | | |
| | | | |
| | | | |
| | | | |

# Table of Contents

# Executive Summary

As next-generation information technologies like AI and big data continue to drive digital and intelligent transformation in industries, China Mobile is capitalizing on the digital economy by developing a cutting-edge "connection + computing + capability" information service system. The system leverages 5G, computing networks, and Ability as a Service (AaaS), expected to promote the construction of intelligent networks.

Intelligent wireless networks are crucial for network intelligence. Their capability evolution not only determines the iterative upgrades of the communications network architecture but also affects the assurance and improvement of services such as network energy saving and air interface transmission. As such, AI must be introduced and integrated into mobile communications networks, encompassing their architectures, functions, and key technologies, to accelerate the network evolution towards intelligence through smart profiling, prediction, optimization, and decision-making. This will facilitate the digital and intelligent transformation of both economy and society.

# Abbreviations

| Abbreviation | Explanation |
|---|---|
| AaaS | Ability as a Service |
| AI | Artificial Intelligence |
| AMC | Adaptive Modulation and Coding |
| BLER | Block Error Ratio |
| CCSA | China Communications Standards Association |
| CQI | Channel Quality Indicator |
| DOA | Direction Of Arrival |
| FTP | File Transfer Protocol |
| KPI | Key Performance Indicator |
| MCS | Modulation and Coding Scheme |
| MTTR | mean time to restoration |
| MU-MIMO | Multi-User Multiple-Input Multiple-Output |
| NE | network element |
| OMC | operation and maintenance center |
| OSS | operations support system |
| PCI | physical cell identifier |
| PDCP | Packet Data Convergence Protocol |
| PDSCH | Physical Downlink Shared Channel |
| PM Counter | Performance Management counter |
| PRB | Physical Resource Block |
| RF | radio frequency |
| RI | Rank Indicator |
| RRC | Radio Resource Control |
| SE | Spectral Efficiency |
| UE | User Equipment |
| | |
| | |
| | |

# Introduction

Next-generation information technologies like AI and big data are now driving digital and intelligent transformation across various industries and are key to improving network efficiency and capabilities. For China Mobile, the construction of intelligent 5G-A wireless networks is essential. It serves as the foundation for China Mobile's vision of network intelligence and drive continuous advancements in network capabilities.

However, the industry is still in the application exploration and capability building phase in terms of 5G-A wireless network intelligence. The overall R&D investment in intelligent products lags behind expectations. Against this backdrop, China Mobile proposes a comprehensive system for evaluating how intelligent a 5G-A wireless network is. By promoting evaluation-driven exploration and investment, we aim to provide industry stakeholders with a reference and guidance for planning, designing, evaluating, and verifying technologies, solutions, and products related to 5G-A wireless network intelligence.

This white paper illustrates the evaluation system and showcases practical application scenarios. It is prepared by China Mobile and co-authored by Huawei, ZTE and VIAVI Solutions.

# 1 Wireless Network Intelligence Evaluation System

The convergence of AI technologies with hardware, software, systems, and processes in communications networks gives rise to network intelligence. AI technologies make the operation of communications networks intelligent, enable agile service innovations, and usher in networks with native AI. This results in higher quality and efficiency of communications networks and supports the digital intelligence of industries. As a subfield and infrastructure of network intelligence, wireless network intelligence integrates AI algorithms and computing power with existing 5G and 5G-A network elements (NEs), unlocking the full potential of AI on networks and driving improvements in performance and efficiency.

## 1.1 Evaluation System

With the evolution of wireless AI applications, traditional indicator-based evaluation systems fall short in meeting the diverse requirements of 5G-A intelligence evaluation scenarios. Moreover, designing evaluation indicators for a single purpose proves challenging. Therefore, network intelligence evaluation systems require incremental improvement on a scenario-by-scenario basis. To meet this requirement, we propose the "1-4-1" evaluation architecture, as shown in the following figure.

The architecture is designed based on typical scenarios and incorporates four key elements: evaluation environments, tools, indicators, and interfaces. We standardize common issues in each element, in a bid to foster industry-wide consensus on wireless network intelligence. This architecture also covers both top-down design and methodology, guiding the industry to gradually improve wireless network intelligence evaluation systems.

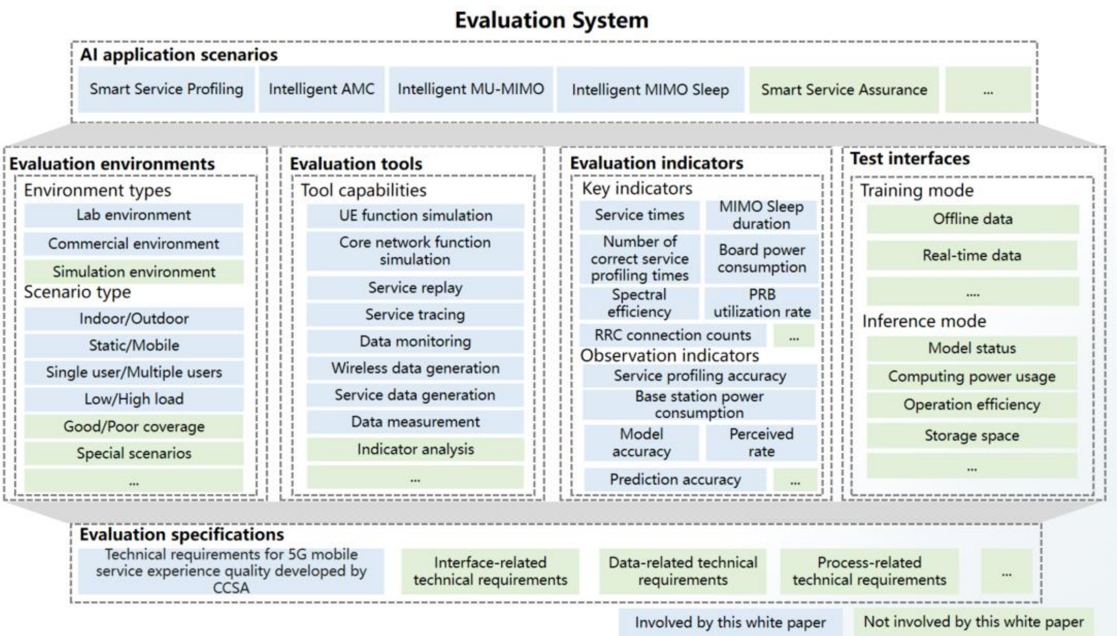


**Figure 1 "1-4-1" evaluation architecture**

In the preceding figure, the blue sections represent content involved in the test cases of this white paper, while green areas denote recommended considerations for future evaluation work, though not explicitly detailed here.

### 1.1.1 Application Scenarios

As wireless networks grow increasingly complex, AI plays a vital role in optimizing their performance. By tailoring AI capabilities to address specific live-network issues, network capabilities can be enhanced and operational efficiency improved. Therefore, scenario-specific case studies are essential to effectively evaluate wireless network intelligence. This white paper covers several typical scenarios.

### 1.1.2 Evaluation Environments

Commercial wireless networks can be classified into distinct categories based on factors like indoor/outdoor coverage, speeds, levels of interference, and capacity requirements. In complex environments, AI with its exceptional fitting and generalization abilities will likely play a pivotal role. When designing an evaluation environment, it is essential to consider the anticipated AI capabilities on the network. For instance, AI-powered prediction can be used in high-speed railway scenarios to optimize resource allocation in high-speed cells that UEs may access. If such a native AI solution is to be evaluated, it must be tested in high-speed railway scenarios to ensure its efficacy. Similarly, evaluating intelligent NEs necessitates considering their intended applications to unlock their full potential in the chosen environment.

### 1.1.3 Evaluation Tools

To effectively evaluate wireless network intelligence, it is essential to define the necessary tools and instruments, including their key functions and requirements. Given the black-box nature of AI, evaluation instruments must offer at least the following functions: simulating core network and UE functions, replaying and tracing services, monitoring data, and generating wireless and service data. While common tools like data monitoring software can be widely applied, differentiated tool functions are also necessary. For instance, to evaluate network performance in service assurance, tools need to provide service recording and replay capabilities, whereas end-to-end coordination test scenarios demand simulation of core network and UE functions. Specific cases are detailed in subsequent sections for further clarification.

### 1.1.4 Evaluation Indicators

To quantify the effects of intelligence, both key indicators and observation indicators must be specified. Key indicators serve as the foundation for observation indicators. Since traditional evaluation methods relying on simple indicators cannot cope with various AI-driven scenarios, observable and diversified evaluation indicators must be formulated based on specific scenario characteristics to evaluate how intelligent wireless networks are. Common wireless network indicators can be used for evaluation, but their application must match scenarios. For instance, shutdown duration serves as a key indicator, while base station energy consumption is an observation one, when assessing intelligent channel shutdown performance. Similarly, the number of RRC connections functions as a key indicator, whereas the user-perceived rate acts as an observation indicator, when evaluating the effectiveness of intelligent MU-MIMO.

This white paper explains these evaluation indicators and demonstrates their practicality through concrete examples. The evaluation methods presented in the following sections are intended solely for verification purposes, and do not represent industry standards or requirements.

### 1.1.5 Test Interfaces

Unlike traditional communications functions, AI applications are functioning while keeping key algorithms concealed. To gauge the effect of a specific AI application, indicators must be presented using standardized interfaces. These interfaces can work with evaluation tools to provide real-time display of the intelligence level of the embedded AI capability.

Test interfaces support both offline output and real-time output. For AI applications in training mode, data to be collected includes offline data such as model initialization parameters and training datasets as well as real-time data that enables NEs to train and update models online when network capabilities permit. For AI applications in inference mode, data to be collected includes the running status, computing power usage, running efficiency, and storage space. When test interfaces are used in future wireless network applications, AI models can be updated promptly and their performance can be monitored in real time.

Due to its nascent stage, the relevant research is not discussed in subsequent sections; nonetheless, it is worth noting that test interfaces hold significant sway over intelligence capability openness and standardization.

### 1.1.6 Evaluation Specifications

To guarantee the scientific and standardized execution of evaluation activities, evaluation specifications must be robust enough to validate the relevance of evaluation environments, tools, indicators, and AI interfaces against industry standards and specifications. Content that can be standardized needs to be identified as well to enhance the standardization of both international and national evaluation systems. For instance, interface definitions, data types, input, and output of network capabilities should be developed based on the requirements of corresponding technical systems, which may include data, interfaces, and processes to facilitate the standardized implementation of evaluation activities.

## 1.2 Application Scope

The native intelligence for wireless networks, enabled by AI, reduces costs and improves efficiency in terms of radio signal processing and radio resource management and also empowers service profiling and evaluation. This intelligence also involves elements for intelligentization such as datasets provided by base stations and wireless operation and maintenance centers (OMCs) to operations support systems (OSSs) and upper-layer AI applications.

The performance evaluation of native intelligence for wireless networks hinges on the deployment entities of AI model inference services. Specifically, when AI model inference is deployed on base stations and wireless OMCs, they must be included as the evaluated objects. When other common or dedicated devices and components serve as training entities, they should be included in the evaluation as well.

# 2 Typical Scenarios

The network can apply intelligence in various ways. For instance, AI-powered analysis helps profile and evaluate network services, while AI-optimized planning enables the network to adjust its settings according to specific scenarios. These diverse applications necessitate an evaluation approach for intelligent networks that moves beyond reliance on a fixed set of indicators. For example, recall and precision rates can be used to evaluate an AI classification algorithm, whereas a convex optimization algorithm must be evaluated in terms of its convergence rate and solution stability. Such an evaluation approach must also define the scenarios in which an algorithm takes effect on the network and what tools are employed to monitor the algorithm's performance. This chapter aims to describe tailored solutions for evaluating how intelligent a wireless network is in specific scenarios, covering both top-down design and methodology. The descriptions in this chapter are organized in line with the "1-4-1" evaluation architecture.

## 2.1 Smart Service Profiling

### 2.1.1 Application Scenarios

Service profiling on the RAN side is a critical technology that gives a significant competitive edge to differentiated service assurance. Without this technology, the network cannot take targeted measures to ensure service quality. This technology also plays a vital role in developing new services, improving user experience, and reducing user complaints.

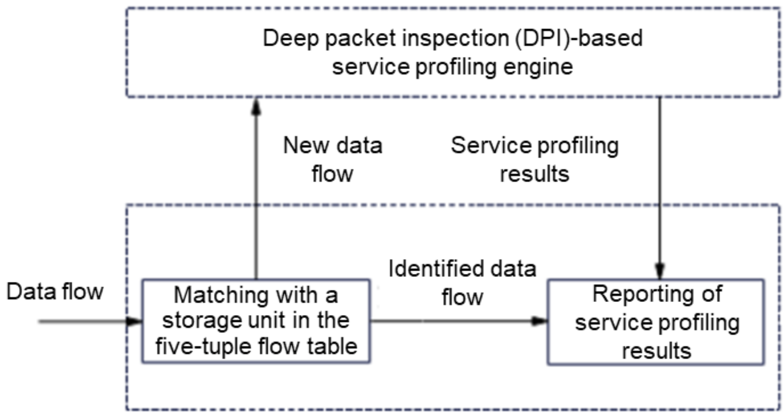


**Figure 2 Smart service profiling**

Service profiling integrates cloud-based training with local inference to create accurate profiles of services. It leverages a database of service features created by an AI model that extracts key user fields and packet features. Specifically, the AI model first collects traffic features of packets (such as certain field values, packet size, and inter-packet intervals). Using semi-supervised learning, the model then learns the correlation between these traffic features and corresponding service types from labeled service samples provided by service dialing tests.

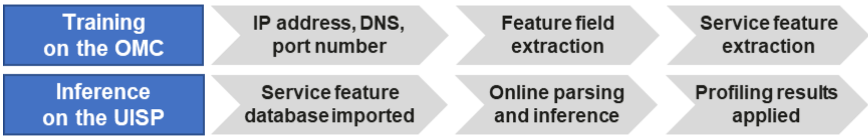


**Figure 3 Process of service profiling**

### 2.1.2 Evaluation Environments

Evaluation of service profiling involves (1) capturing real packets for services on various apps; (2) using an interface protocol tester to replay the live-network services for the network to perform service profiling.

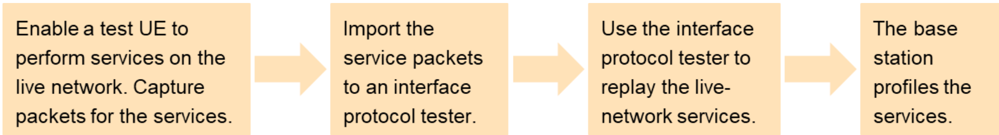


**Figure 4 Process of service profiling evaluation**

Packet capture environment: Use a tool to capture sufficient packets for UEs' real services on the live network (such as short video watching, uplink live streaming, and online gaming). The captured packets will be used for service replay.

Mainly capture packets for encrypted services on the live network, because AI has a strong capability to analyze encrypted services. On the other hand, to ensure the integrity of the evaluation, capture a few packets for unencrypted services as well.

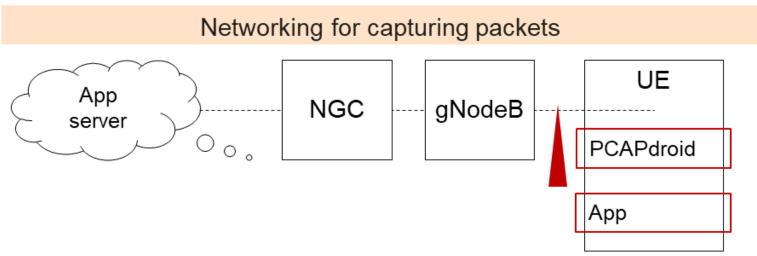


**Figure 5 Networking for capturing packets**

Service replay environment: Use an interface protocol tester to replay the real services on the base station. The base station will receive data from the tester and determine the type of the replayed service.

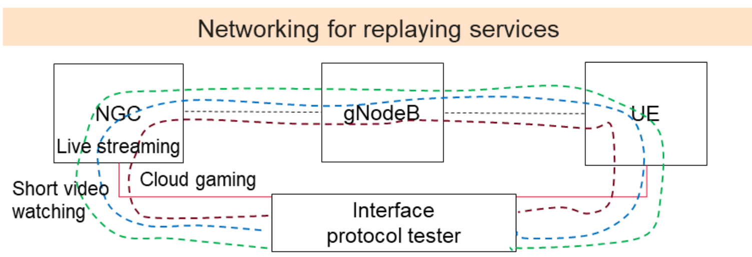


**Figure 6 Networking for replaying services**

During the replay of real services, the captured packets for real services are sent from the interface protocol tester to the core network, base station, and UE in sequence; ACKs from the

UE are sent to the base station, core network, and interface protocol tester in sequence. In this process, the base station profiles the services replayed by the interface protocol tester. The profiling results of the base station are then compared with the services replayed by the tester to check whether they are consistent. This is how the accuracy of service profiling is evaluated.

### 2.1.3 Evaluation Tools

Evaluation of service profiling requires the following tools:

Packet capture tool: It is an app installed on UEs. Once installed and opened, this app will capture packets for all subsequent services. When using this app, mainly capture packets for encrypted services on the live network to evaluate AI's capability to profile encrypted services; capture a few packets for unencrypted services as well to ensure the integrity of the evaluation.

Service replay tool: It is used to replay real services (that is, to replay captured packets for services like short videos, long videos, and mobile games on UEs). The data flows for the replayed services go to the base station in the downlink and to UEs in the uplink.

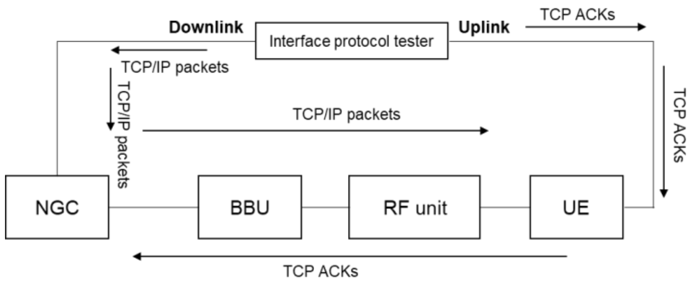


**Figure 7 Networking for and data flows from an interface protocol tester**

Service analysis tool: It is used to record UE logs, analyze UE and service status, and check UEs' network connection, radio environment, physical cell identifiers (PCIs) of served cells, and other information.


**Figure 8 Service analysis tool**

### 2.1.4 Evaluation Indicators

Service profiling is evaluated by comparing two key indicators: the number of times that a UE performs a service and the number of times that the base station correctly profiles the service.

Number of times that the base station correctly profiles a service: This indicator measures how many times the base station correctly profiles a service performed by a specific UE within a specified period (for example, 10 minutes). Whether a profiling result is correct depends on the actual service on the UE.

Number of times that a UE performs a service: This indicator is incremented by one each time a UE performs a service such as short video watching within a specified period (for example, 10 minutes).

Evaluation of service profiling uses the "service profiling accuracy" as the indicator for observation. This indicator is calculated as follows:

$$\text{Service profiling accuracy} = \frac{\sum \text{Number of times that the base station correctly profiles a service}}{\sum \text{Number of times that a UE performs the service}} \text{ x } 100\%$$

In most cases, the average service profiling accuracy (including both encrypted and unencrypted services) should be 95% or higher to ensure sound service assurance and deterministic experience.

### 2.1.5 Evaluation Specifications

There are currently no industry-wide specifications for evaluating smart service profiling. Nevertheless, China Communications Standards Association (CCSA) is moving to develop the technical requirements for 5G mobile service experience quality[1]. This exemplifies that user experience and sustainable network development are taking center stage in building, optimizing, and operating 5G mobile networks. In the future, it is crucial to continue researching the evaluation of smart service profiling and establish widely accepted specifications across the telecommunications industry.

## 2.2 Intelligent AMC

### 2.2.1 Application Scenarios

The wireless channel is influenced by complex physical environmental factors, such as path loss, multi-path fading, and doppler frequency offset, resulting in time-frequency dual fast-varying characteristics. This causes rapid fluctuations in the channel quality between a base station and a terminal within millisecond-level time slots. Furthermore, complex and irregular interference signals exacerbate performance variability across time slots, where uniform scheduling strategies may waste spectral resources in high-quality slots and lead to significant packet loss in low-quality slots, causing sharp performance degradation.

In current networks, notable inter-slot performance disparities occur in scenarios such as atmospheric ducting causing uneven slot interference, different pilot structures between S-slots and D-slots in high-speed rail settings, and interference in specific overlapping slots under macro-micro networking. For example, in atmospheric ducting scenarios, demodulation differences between time slots under traditional scheduling strategies result in low-performing slots dragging down high-performing ones, thereby impacting system performance and user experience. Therefore, the key to addressing rate instability in such scenarios is the intelligent and adaptive adjustment of scheduling strategies to swiftly and efficiently match channel variations.

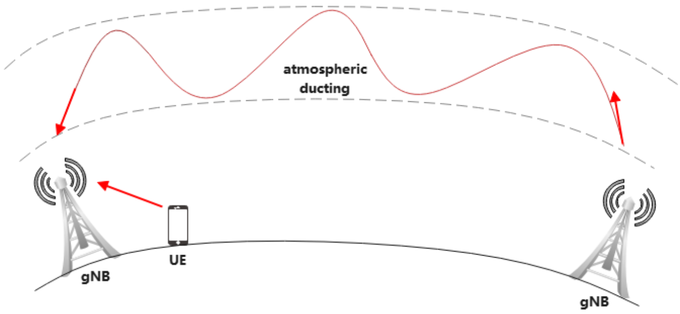


**Figure 9 Atmospheric duct interference diagram**

The main concept of Intelligent AMC (Adaptive Modulation and Coding) is using clustering algorithms to group different time slots for the same terminal based on channel characteristics, applying homogeneous scheduling strategies within clusters and heterogeneous strategies between clusters. This enables differentiated scheduling according to high-order channel features of each slot, thereby fully exploiting the performance potential of both high- and low-quality slots. To further reduce time-frequency overhead from exploratory processes and accelerate the convergence of intra-cluster scheduling strategies, big data is used to train a scheduling optimization network, integrating historical feature data to promptly adjust existing strategies and quickly converge to the optimal scheduling position. Combining these two approaches allows for stable scheduling across time slots in rapidly varying channel/interference scenarios, reducing the impact of anomalous slots on scheduling stability and maximizing spectral efficiency.

The design of the clustering algorithm is central to this intelligent network element. The clustering algorithm categorizes sample points based on the similarity of selected wireless features, ensuring high intra-cluster cohesion and low inter-cluster coupling. This approach effectively differentiates high-quality from low-quality channels, mitigating the impact of environmental variations on scheduling strategies, maintaining stability in the system’s Block Error Rate (BLER), and maximizing the capacity potential of each time slot.

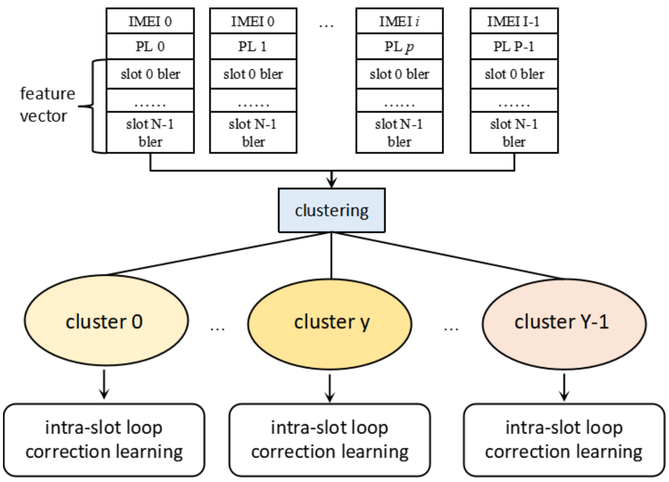


**Figure 10 User clustering flowchart**

### 2.2.2 Evaluation Environments

This application focuses on the use of clustering algorithms to apply differentiated scheduling strategies to high- and low-quality time slots, thus leveraging the performance gains from selectively releasing capabilities across time slots. This capability shows significant benefits in scenarios with slot imbalance. Accordingly, during testing, this example constructs scenarios where time slot performance differs markedly, using BLER and spectral efficiency metrics to demonstrate the reliability and effectiveness of the algorithm.

In laboratory testing, this example constructs a single-user, indoor wired environment with controlled interference, where interference is artificially introduced into fixed time slots to generate performance variations across time slots. For commercial site testing, sites in a commercial network are selected where there are at least 10 5G users and a minimum of 2 types of commercial terminals. Points with substantial slot scheduling variation are selected, specifically those where the BLER difference between high- and low-quality slots exceeds 20%. At these points, a UE is connected to perform single-user FTP traffic.

### 2.2.3 Evaluation Tools

In laboratory testing, a single UE performs full-load FTP traffic, using a vector signal generator as a testing tool. Interference is introduced at fixed time slots, with the form of interference unrestricted, as long as it transmits in a fixed downlink slot. The interference intensity is set between 2-5 dB, with testing conducted for one hour.

In commercial site testing, a network packet-filling tool is used to conduct FTP traffic on commercial terminals. Test points with notable differences in BLER between high- and low-quality slots are selected. Downlink spectral efficiency at these points is compared across multiple time granularity—hourly, daily, and weekly—with and without the intelligent AMC function enabled. In the above tests, the core evaluation tools primarily include two categories: wireless data generation and service data generation. During testing, a vector signal generator and a network packet-filling tool are respectively used to generate interference and service signals.

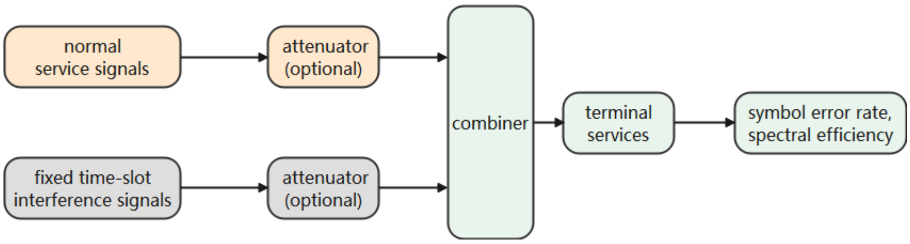


**Figure 11 Laboratory test flowchart**

### 2.2.4 Evaluation Indicators

During testing, the slot-level BLER and spectral efficiency were recorded before and after enabling the intelligent AMC function. Both metrics are accessible from the network side, with slot-level BLER averaged by slot and spectral efficiency averaged overall. After enabling the intelligent AMC function, a 15-minute convergence period was allowed, followed by an additional hour of testing to record the slot-level BLER and spectral efficiency under this setting.

$$\text{Block Error Rate} = \frac{\text{Code Blocks with Demodulation Errors}}{\text{All Transmitted Code Blocks}} * 100\%$$

$$\text{spectral efficiency} = \frac{\text{Total TBSize of Normally Transmitted Data}}{\text{Total Downlink PRBs}}$$

In the laboratory scenario, the BLER of interfered slots was initially very high before enabling the function, with specific values depending on interference intensity; in extreme cases, the interfered slot BLER could exceed 90%. After enabling the function, the BLER across all slots became more balanced, stabilizing around 10%, indicating that the intelligent AMC function was effectively scheduling high-order resources to high-quality slots and lower-order resources to low-quality slots based on slot quality differences, thereby fully leveraging each slot’s resource advantage. As a result, cell spectral efficiency improved significantly, with gains of 3-10% across various interference scenarios.

In commercial testing scenarios, spectral efficiency increased from 3.074 bps/kHz before enabling the function to 3.280 bps/kHz after, reflecting a 6.7% gain. Additionally, cell downlink BLER decreased from 13.11% to 11.06%, a reduction of 2%, further demonstrating the effectiveness and reliability of this intelligent algorithm.

Currently, due to limitations in evaluation methods, assessments focus primarily on system performance. In the future, a more comprehensive evaluation could include the AI algorithm’s intrinsic performance (such as performance under multiple interference sources), response speed (e.g., convergence time for algorithm effectiveness), and memory usage.

### 2.2.5 Evaluation Specifications

This application primarily relies on the base station to select the most suitable transmission rate based on terminal status information, utilizing relevant standardized interfaces for user status awareness, such as the A/N feedback interface. However, the evaluation does not currently incorporate relevant standards and specifications. In the future, if intelligent scheduling strategy optimization is jointly implemented by the base station and terminal based on channel state matching, it will likely involve more detailed state information reporting from the terminal side and associated standardization efforts.

## 2.3 Intelligent MU-MIMO

### 2.3.1 Application Scenarios

MU-MIMO is a technology that leverages multiple antennas at both transmitter and receiver ends to gain diversity, fully utilizing spatial resources to simultaneously transmit and receive data for multiple UEs within the same frequency band, thereby enhancing cell throughput. Traditional MU-MIMO calculates correlations by iterating over all terminals, which results in high computational complexity, long iteration cycles, and limited spatial multiplexing performance. Intelligent MU-MIMO addresses these issues by introducing intelligent models in terminal pairing selection and post-pairing AMC convergence.

The intelligent MU model primarily operates by extracting spatial and channel quality features from users, using historical data to train and build user profiles. Through clustering and elite ant colony algorithms, the model accurately pairs users with identifiable spatial characteristics in high-capacity environments. The spatial multiplexing strategy model and the MCS translation and tracking model are continuously updated through online training. In the spatial multiplexing strategy model, grid partitioning is performed in the cell's angular domain rather than by geographic location, with the optimal pairing inference model trained using MR data. The MCS translation and tracking model is trained with user profiles that encompass DOA spatial features, channel quality indicators such as CQI and RI, and historical scheduling information, including transmission modes and the number of spatial streams.

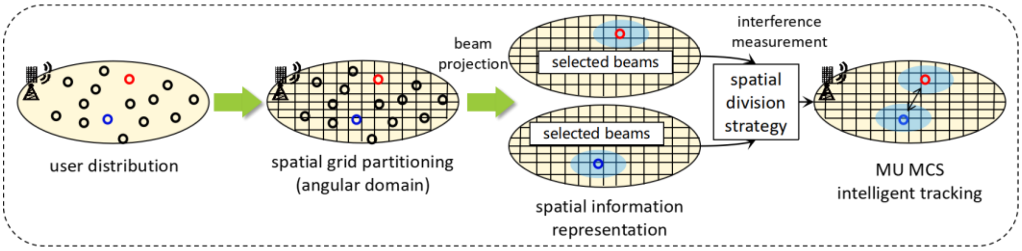


**Figure 12 Intelligent MU-MIMO flowchart**

### 2.3.2 Evaluation Environments

This application primarily focuses on the benefits of spatial grid partitioning and clustering algorithms in spatial multiplexing scenarios. The gains of the AI algorithm are reflected in its efficiency and accuracy compared to traditional correlation-based MU algorithms, which are most evident in high-traffic scenarios. Therefore, to construct an optimal gain scenario, it is necessary to ensure that the cell has over 200 RRC users within its coverage area.

In laboratory testing, this example constructs an indoor wired multi-user environment, simulating MU traffic scenarios with over 200 RRC connections. In commercial site testing, sites with at least 200 5G users and 10 types of commercial terminals are selected to validate the effectiveness of the intelligent MU-MIMO application in high-capacity environments.

### 2.3.3 Evaluation Tools

The evaluation scenarios for this application are high-capacity business environments. Performance gains are assessed by comparing KPI metrics with and without the AI algorithm under normal traffic conditions. In laboratory testing, a multi-user terminal simulator is used to emulate a high-capacity environment, ensuring the RRC connection count remains above 200. In commercial site testing, 200 RRC users conduct various services such as FTP, video, and voice, with downlink spectral efficiency evaluated under different time granularities with and without the intelligent MU-MIMO application enabled.

Key evaluation tools include terminal function simulators and service data generators, with the multi-user terminal simulator and network packet-filling tool used to simulate multi-user traffic scenarios and generate service signals in realistic testing environments.

### 2.3.4 Evaluation Indicators

This application focuses primarily on optimizing the multi-user spatial multiplexing pairing strategy, with significant performance gains observed in cells with extremely high traffic volumes. In this test, a commercial site with real-world business scenarios is selected for model training and deployment testing, choosing cells with more than 300 RRC connections during peak hours. To evaluate the algorithm's effectiveness, key metrics such as downlink PDCP layer traffic, downlink PDCP spectral efficiency, and downlink UE-perceived rate are compared between traditional MU operation and AI-enhanced MU operation, with KPI metrics monitored from the base station network management, including:

Average number of connected RRC users, Downlink spectral efficiency, Downlink UE-perceived rate (Mbps), Downlink MU-SE and SU-SE, Cell-level downlink PDCP traffic (GB), Average number of scheduling layers per cell, Downlink average MCS, Proportion of downlink MU-MIMO spatial multiplexing slots, PRB utilization rate of the downlink shared channel, BLER for PDSCH spatial multiplexed groups, BLER for non-spatially multiplexed PDSCH groups, PRB proportion of MIMO-paired downlink service channels, Average CQI, Average layer count for downlink MU-MIMO scheduling.

$$\text{Perceived Rate} = \frac{\text{Total TBSize of Correctly Transmitted Data}}{\text{Duration of the Given Period}}$$

Among, the improvements in downlink spectral efficiency and UE-perceived rate directly reflect the spectral efficiency gains from the intelligent model's optimization of MU pairing and closed-loop inference.

Performance is assessed across downlink PDCP layer traffic, downlink PDCP spectral efficiency, and downlink UE-perceived rate, comparing RRC user counts and CQI averages before and after enabling the intelligent MU algorithm, ensuring that user count and CQI averages are consistent. In the baseline MU pairing algorithm, spectral efficiency tends to decrease as RRC connections increase. However, with the intelligent MU algorithm enabled, downlink spectral efficiency improves across varying RRC connection counts, particularly in high RRC scenarios. With 300-350 RRC connections, downlink spectral efficiency gains peak at 44.25%, fully validating the effectiveness of this application.

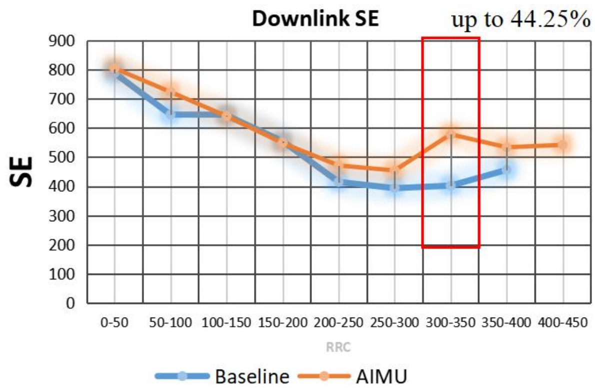


**Figure 13 Intelligent MU-MIMO test results**

### 2.3.5 Evaluation Specifications

In this application, the base station determines multi-user or single-user scheduling based on the reported status from multiple users. Relevant standard interfaces, such as DOA spatial features, CQI, and RI, are used for user status awareness, though the evaluation does not cover specific standards or protocols. In the future, if the base station and terminal jointly implement multi-user/single-user scheduling, additional interactive interfaces between the base station and terminal sides will be required.

## 2.4 Intelligent MIMO Sleep

### 2.4.1 Application Scenarios

The intelligent MIMO Sleep feature utilizes AI technology to provide more accurate predictions of future traffic patterns by learning from historical traffic data, compared to traditional rule-based power-saving strategies. This allows for more opportunities for sites to enter power-saving modes, extending the duration of MIMO Sleep and achieving greater energy savings.

The application supports online learning and prediction. It can initially operate in detect mode, where it learns over a period of time before activating the feature, which helps improve prediction accuracy. It also supports directly activating the intelligent MIMO Sleep mode. The AI algorithm undergoes a learning period, and once the accuracy reaches a certain level, the feature is gradually enabled, minimizing the impact on the live network.

Traditional rule-based MIMO Sleep primarily rely on parameter and time window configurations to enable or disable the feature. In contrast, intelligent MIMO Sleep uses AI technology to predict traffic and enters energy-saving mode when thresholds are met, optimizing energy consumption around the clock.

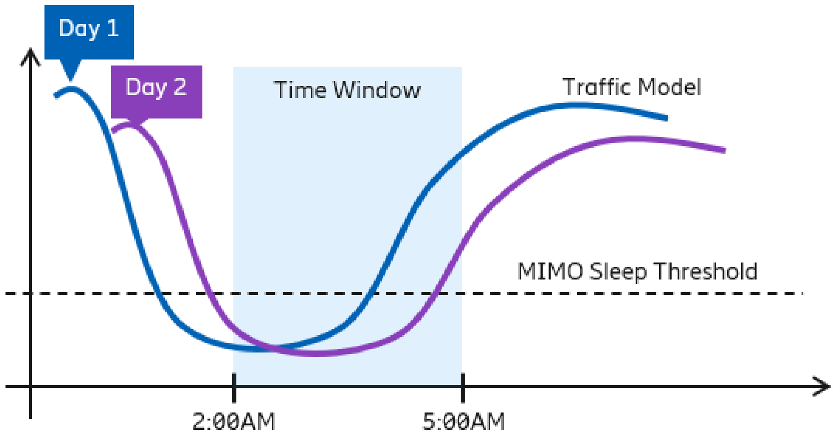


**Figure 14 Traditional rule-based MIMO Sleep**

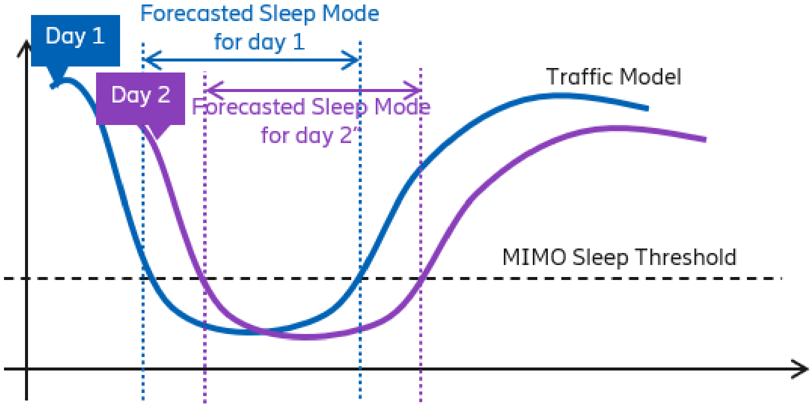


**Figure 15 Intelligent MIMO Sleep**

The intelligent MIMO Sleep is applicable to all scenarios in the live network. It uses AI models to predict and monitor network traffic 24/7. Once the network enters a low-traffic model, the site will automatically enter power-saving mode. It also provides real-time monitoring of the network, and if actual traffic exceeds predicted traffic, the site will automatically exit power-saving mode.

By incorporating AI technology and learning from historical data, the intelligent MIMO Sleep feature can predict future traffic patterns, thereby extending the duration of power savings. The model predicts the PRB usage for the next period every 5 minutes. If the predicted PRB utilization is below a threshold, the entire prediction interval is marked as a "candidate" period for MIMO Sleep. Every minute, the predicted PRB utilization is compared with the actual PRB usage, which helps to quickly deactivate the feature in case of prediction failure, minimizing the impact on the live network.

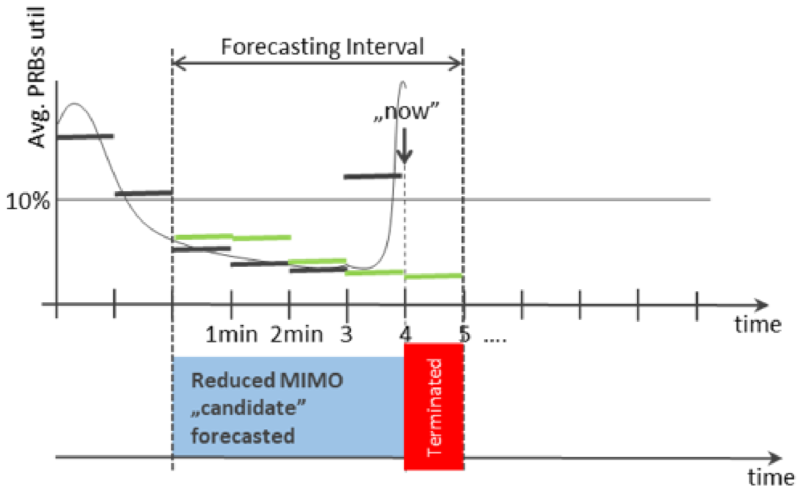


**Figure 16 Prediction of Future Traffic Model by Intelligent MIMO Sleep Function**

## 2.4.2 Evaluation Environments

The intelligent MIMO Sleep technology is suitable for any high, medium, and low-load scenarios in the live network.

High-traffic sites: These are selected in city centers or commercial areas, where traffic volume is usually high on both weekdays and weekends.

Medium-traffic sites: These are selected in urban residential areas, where traffic volume fluctuates between weekdays and weekends.

Low-traffic sites: These are selected in suburban or rural areas, where traffic volume is generally low during most periods but may increase during holidays.

The testing periods cover 24 hours a day from Monday to Friday for weekdays, 24 hours a day on Saturday and Sunday for weekends, and 24 hours a day during legal holidays.

By collecting traffic patterns from different scenarios at different times, including weekdays, weekends, and holidays, the model can be accurately evaluated.

### 2.4.3 Evaluation Tools

The intelligent MIMO Sleep feature is based on AI models to predict traffic on the live network, eliminating the need for additional evaluation tools. The assessment can be completed using existing metrics.

### 2.4.4 Evaluation Indicators

The intelligent MIMO Sleep technology utilizes a configured PRB utilization threshold to automatically enable and disable the feature by comparing predicted and actual traffic volumes.

When both the actual PRB utilization and the predicted PRB utilization are above the set threshold, the feature is not activated, meaning the MIMO Sleep mode is not entered.

When both the actual PRB utilization and the predicted PRB utilization are below the set threshold, the feature is activated, meaning the MIMO Sleep mode is entered.

When the actual PRB utilization is higher than the predicted PRB utilization, the feature will automatically deactivate, exiting the MIMO Sleep mode.

The feature is enabled by comparing the predicted PRB utilization with the actual PRB utilization. When the PRB utilization is below the set threshold, the feature is activated. When it is above the set threshold, the MIMO Sleep mode is exited.

By observing the live network KPI, calculate the accuracy of the model using the following formula:

$$\text{Actual AI model accuracy} = \frac{\text{Actual duration of MIMO Sleep duration}}{\text{Predicted duration of entering sleep mode}} * 100\%$$

Collect PM Counter data for one hour and observe the periods within that hour where the PRB utilization is below the configured threshold. The model's prediction accuracy is evaluated by comparing the actual duration of MIMO Sleep with the expected duration predicted by the model.

The intelligent MIMO Sleep technology can accurately predict traffic under the network's varying traffic models, achieving a prediction accuracy of over 90%. By using AI models for traffic prediction, sites have more opportunities for power savings. Compared to traditional MIMO Sleep methods, the power-saving ratio is improved by more than 5%. This means that intelligent MIMO Sleep technology can further enhance energy efficiency while ensuring communication quality.

### 2.4.5 Evaluation Specifications

Intelligent MIMO Sleep is an emerging AI-native technology, and the metrics it involves (such as predicted MIMO Sleep opportunities and the number of times MIMO Sleep is terminated) have not yet been included in any standard specifications. These metrics involve reporting and interaction between base stations and Network Manager. If this feature is introduced in the future, new metrics will need to be added to the relevant interfaces in the network management standard. These could include the number of predicted opportunities to enter MIMO Sleep, the number of times the intelligent MIMO Sleep feature is enabled to enter or remain in sleep mode, and the number of times the intelligent MIMO Sleep feature terminates sleep mode due to unpredictable traffic increases.

# 3 Development Trends

The escalating level of wireless intelligence is anticipated to empower more sophisticated and widespread intelligent applications in the years to come. Future applications will primarily enhance basic wireless communications services in the near and mid-term. In the long run, edge AI computing could become a major trend. This chapter delves into how AI transforms wireless communications in terms of experience assurance, energy saving, and O&M.

Experience assurance: Intelligence is driving the network to evolve from delivering differentiated experience to ensuring deterministic experience. This shift requires the network to provide end-to-end service assurance, which builds on two key capabilities: intelligent prediction and intelligent assurance. Intelligent prediction enables the network to predict its capability to provision services, specifically by forecasting how many VIP users can be guaranteed based on predicted traffic, cell resource load, and interference. Intelligent assurance ensures that SLA-compliant experience is delivered for VIP users and key services, protects the experience of background UEs, and evaluates service experience. With intelligent assurance, resource allocation between different UE and service levels is optimized through setting various targets in terms of experience, resources, and energy consumption. Telecom carriers must possess these intelligence capabilities to deliver deterministic experience on their networks and achieve business success.

Against this background, evaluation of experience assurance will place greater emphasis on deterministic experience. Take assurance of short video watching as an example. Traditional evaluation focuses on how rates and delays change after the service is assured, with evaluation models, tools, and environments all centered around these changes. In contrast, future testers should first establish the conditions for 1080p short videos, so as to set standards for the deterministic experience of this service. With standards in place, testers can then design models, tools, and environments that aim to evaluate whether the standards can be met. The same principle applies to evaluating other services such as live streaming, gaming, and web browsing.

Energy saving: Balancing energy saving with optimal user experience is a major topic of interest that continues to be a pressing concern. Achieving this balance requires intelligence. Reinforcement learning, for example, can help set energy saving parameters and features tailored to each site's needs, allowing each site to follow a unique, optimal policy for energy saving. Moreover, intent-driven energy saving is maturing due to advancements in intelligence. This approach enables the maximum energy efficiency while meeting multiple intents (targets). Once users define their expected experience, energy consumption, delay, and other intents for different scenarios, the system can automatically complete end-to-end energy saving actions, including translating intents, generating and delivering policies, and evaluating effects.

To keep pace with the changes brought about by intelligence, evaluation of energy saving should no longer focus solely on how much energy is saved. Instead, a new approach must be adopted for intent-driven energy saving, evaluating how well a system adapts, how accurately it translates user intents, and whether it satisfies user intents in terms of experience, delay, and other aspects while saving energy. This shift requires the development of standardized evaluation methods and tools.

O&M: Unmanned network O&M is now on the horizon. One example of it is a smart troubleshooting assistant that leverages cutting-edge technologies such as large models. Unlike in traditional O&M, where foreground and background engineers have to interact with each other, engineers will be able to interact with a robot assistant through text and voice to get various information such as equipment status, alarms and fault diagnosis, configurations, and network topology. This intelligent O&M will more efficiently assist agent maintenance personnel in troubleshooting. Another application of intelligence in O&M is the use of components like intelligent optical modules at base stations. These components help base stations better learn the topology of surrounding dummy devices and their status changes, so that faults can be more accurately located. As a result, it is possible to resolve issues without having to alter the network, or by visiting sites only once to make the necessary changes to the network.

Evaluation of intelligent O&M also differs significantly from traditional evaluation in terms of both the objects being evaluated and the methods used. As intelligent O&M uses large models and other new technologies, the interface for evaluation will be expressed in humans' natural languages. This will introduce a realm of uncertainties that must be taken into account when constructing evaluation data and analyzing results. An ideal evaluation of intelligent O&M must be centered around customers, particularly customers' service scenarios and work flows. While recall and precision rates remain important indicators, they should not be the sole focus. More weight should be given to whether the troubleshooting efficiency is improved (as evidenced by a lower mean time to restoration (MTTR), for example), how efficient information query is, and how well customers accept diagnostic results provided during troubleshooting.

As 5G-A wireless networks become increasingly intelligent, evaluation tools are shifting their focus from evaluation for AI to using AI for evaluation. Current evaluation tools can be used to create a virtual environment. Using these tools, users may add physical entities such as buildings, base stations, and moving cars to accurately simulate a scalable scenario where an access network could be deployed. These simulations can run on either a digital synthetic map or a real geographic map, mimicking various processes like channel behaviors, network slicing, RAN protocol stacks, energy saving, and KPI generation. The generated datasets can then be used for AI or machine learning, ultimately helping optimize network configuration and policies. This is what evaluation for AI means. Nevertheless, evaluation tools need to go beyond simulating wireless networks to creating digital twins. They must support AI natively as wireless network management is becoming more intelligent and real-world wireless networks are growing in complexity. Digital twinning with built-in AI will more efficiently create virtual representations that are more scalable. Examples of this technology include using AI-powered generative models to simulate access-network scenarios, live-network traffic, and PKI model parameters, automating networks through AI, and replacing complex RF models with AI models.

In conclusion, in the years to come, the integration of advanced intelligence into networks is likely to grant networks even greater capabilities. Digital twins will enable the creation of highly accurate digital replicas of sites and networks so as to reflect realities, predict future developments, and optimize networks. These capabilities can be leveraged for a wide range of applications, including dummy device management, fault prediction and prevention, and rapid network optimization.

# 4 Summary and Prospects

As network intelligence continues to iterate and evolve, its applications are rapidly expanding, unlocking boundless opportunities for digital and intelligent transformation across industries. The wide application of intelligence in every facet of society will revolutionize the way people live and work. This white paper introduces the "1-4-1" architecture, a pioneering framework for evaluating how intelligent a 5G-A wireless network is, and provides evaluation systems for typical scenarios that outshine traditional indicator-based systems. Therefore, this document can provide guidance and theoretical basis for intelligent evaluation in the industry and facilitate the industry consensus on system improvement.

To harness the transformative power of AI, a collaborative ecosystem must be fostered among industry stakeholders, universities, researchers, and applications. We need to push the boundaries of scientific discovery and translate cutting-edge advancements into practical solutions to enrich human life.

# 5 References

[1] Technical requirements for 5G mobile service experience quality developed by CCSA