https://doi.org/10.26599/AIR.2022.9150003 CAAI Artificial Intelligence Research

Next Decade of Telecommunications Artificial Intelligence

Ye Ouyang[1] ✉, Lilei Wang[1], Aidong Yang[1], Tongqing Gao[2], Leping Wei[3], and Yaqin Zhang[4]

ABSTRACT

It has been an exciting journey since the mobile communications and artificial intelligence (AI) were conceived in 1983 and 1956. While both fields evolved independently and profoundly changed communications and computing industries, the rapid convergence of 5th generation mobile communication technology (5G) and AI is beginning to significantly transform the core communication infrastructure, network management, and vertical applications. The paper first outlined the individual roadmaps of mobile communications and AI in the early stage, with a concentration to review the era from 3rd generation mobile communication technology (3G) to 5G when AI and mobile communications started to converge. With regard to telecommunications AI, the progress of AI in the ecosystem of mobile communications was further introduced in detail, including network infrastructure, network operation and management, business operation and management, intelligent applications towards business supporting system (BSS) & operation supporting system (OSS) convergence, verticals and private networks, etc. Then the classifications of AI in telecommunication ecosystems were summarized along with its evolution paths specified by various international telecommunications standardization organizations. Towards the next decade, the prospective roadmap of telecommunications AI was forecasted. In line with 3rd generation partnership project (3GPP) and International Telecommunication Union Radiocommunication Sector (ITU-R) timeline of 5G & 6th generation mobile communication technology (6G), the network intelligence following 3GPP and open radio access network (O-RAN) routes, experience and intent-based network management and operation, network AI signaling system, intelligent middle-office based BSS, intelligent customer experience management and policy control driven by BSS & OSS convergence, evolution from service level agreement (SLA) to experience level agreement (ELA), and intelligent private network for verticals were further explored. The paper is concluded with the vision that AI will reshape the future beyond 5G (B5G)/6G landscape, and we need pivot our research and development (R&D), standardizations, and ecosystem to fully take the unprecedented opportunities.

KEYWORDS

artificial intelligence (AI); mobile communication; 5th generation (5G); general purpose technology (GPT); network intelligence; intent-based network; network AI signaling system

The commercial development of mobile communication technology has lasted for 37 years. From Bell Laboratory and Motorola's large-scale commercialization of the first generation (1G) of analog voice communications technology—advanced mobile phone system (AMPS) in October 1983 as the origin, to the world's mainstream second generation (2G) of mobile communications technology—global system for mobile communications (GSM) which achieved the full digital voice telephony in 1991, from the third generation (3G) universal mobile telecommunications system (UMTS) in 2001, which supports voice and mobile data services, to the fourth generation (4G) of mobile communications technology—long term evolution (LTE), which has been commercially available globally since 2008, supports all-internet protocol (all-IP) high-definition voice and high-speed mobile data services, finally the 5th generation (5G) technology became commercially available worldwide in 2018[1].

1 Mobile Communication and AI

Over the past 30 years, mobile communications have achieved development and evolution of technology and ecology from analog to digital, from voice to voice and data services to pay equal attention, from circuit switching to IP, and from an enclosed communication ecosystem to an enabling vertical industry. In the early stage of mobile communication development, especially from 1G to 3G, the ecosystem of mobile communication network and business is still constantly being constructed. Till the 4G ecosystem basically realized all-IP network system, supported voice and data services, and tried to start the vertical industry empowerment, the industry began to put forward the demand and development concept of automation and intelligence of mobile communication network. As mobile communication network becomes increasingly complex and communication service ecology becomes increasingly diversified, communication network infrastructure and service system need to confront many complex scenarios such as the extremely complex wireless environment that cannot be simulated with physical models, exponential IP switching and routing control options, proactive network support and business assurance, "one customer one policy" and "one moment one policy" network personalized service, etc., far exceeding the processing and management capabilities of traditional manual rule pre-definition and

1 AsiaInfo Technologies (China) Ltd., Beijing 100193, China
2 China Mobile Communications Group Co., Ltd., Beijing 100032, China
3 China Telecom Group Corporation, Beijing 100033, China
4 Institute for AI Industry Research, Tsinghua University, Beijing 100084, China
Address correspondence to Ye Ouyang, ye.ouyang@asiainfo.com

execution. Therefore, current communication system needs a set of automated and intelligent systems and methods to guarantee the operation and development of network and service.

From the commercialization of 3G in 2001 to the widespread commercialization of 5G in 2020, mobile Internet and data services have flourished in these 20 years. Massive amounts of big data are generated in the communication ecosystem, providing a natural and high-quality data source for the development and application of AI in the communication field. In 2006, Hinton et al.[2] proposed deep learning, which marked the rise of the third wave of AI development. Supervised learning, unsupervised learning, reinforcement learning in traditional machine learning, and the neural network in AI have been applied in various scenarios in the communication field in the form of deep learning. By searching the IEEE Xplore database with keywords such as "artificial intelligence", "machine learning", and "deep learning" for academic papers in the field of communication AI, we found that the number of papers from 2006 to the present were 6.42 times as high as the 15 years before 2006. Consequently, since the third development wave of AI in 2006, the integrated application of AI and mobile communication industry has entered a rapid development stage.

General purpose technology (GPT) usually refers to those technologies that can influence global or national economies. GPT is promising to significantly change the society by influencing existing economy and social structure[3–5]. Economists Lipsey et al.[6] defined 24 technologies such as AI as the GPT back in 2005. Since 2018, governments and academic organizations around the world have gradually begun to treat 5G as a new generation of GPT[6–10]. The distinctive feature of GPT are its technological diffusion and empowerment to various industries and its productivity boost for research and development (R&D) and innovation in vertical industries[4]. 5G and AI have such characteristics obviously. Therefore, 5G and AI are generally regarded by countries and industry as the latest set of general-purpose technologies to be adopted in the 21st century.

As 5G was gradually commercialized in 2018, many literatures[1, 11, 12] explored the application of AI in 5G in the form of survey or empirical research, most of which did an overview of AI application cases from 5G physical layer, medium access control (MAC) layer, network layer and application layer, or conducted simulation or data analysis for field experiment or empirical studies with a certain research question. However, the industry still lacks and needs an overall review and forward-looking perspective on the integration and development of mobile communication and AI technology from the perspective of 5G and beyond 5G (B5G) ecosystem development. We hope to use 5G and AI technology as a set of general purpose technology, and take their integration and application development as the main line, and provide a systematic overview of the "intelligence" and "integration" of AI in the current 5G international technology standards, 5G networks and service ecosystems, and then make a forward-looking discussion on the convergence and evolution of mobile communication and AI technologies in the next 10 years.

2 Development Roadmap of Mobile Communication and AI

Mobile communication technology and AI have distinct and independent development roadmap in their respective early stages. In the development of mobile communication technology from 2G to 5G, the industry basically evolves with 3rd generation partnership project (3GPP) as a main line of de facto technical standards, with European Telecommunications Standards Institute (ETSI), International Telecommunication Union (ITU), Open RAN Alliance (O-RAN), and other technical standards as a sideline. Since 2008, 3GPP has gradually introduced the AI concept into the technology standards of the mobile communication network, with self-organizing networks (SON) technology as a significant mark.

2.1 Historical development roadmap of AI

The prototype of AI technology first appeared in 1956, and the term "artificial intelligence" was raised at the Dartmouth Conference[13]. In the same year, Samuel[14] proposed the theory of machine learning, as shown in Fig. 1. In the mid-1970 s, the bionics-based research schools gradually became popular, and technologies such as expert systems and neural networks achieved rapid development[15]. Since then, people have attempted to study general AI programs, but encountered serious obstacles, stagnated and entered a "cold winter" period of development[16]. In 1997, the success of "Deep Blue" put the development of AI on the agenda again. With the increase of computing power and the massive data brought by the popularization of the Internet, the bottleneck of AI has been broken, providing the possibility for the development for deep learning and reinforcement learning based on big data. In the early 21st century, the development of AI technology has made important progress from "perception" to "cognition", especially in deep learning technologies such as speech processing, text analysis, and video processing. In 2012, Hinton[17] published a well-designed convolutional network AlexNet, adding rectified linear units (ReLU) and dropout processing method to the traditional convolutional network, and expanding the network structure to a larger scale, which greatly reduces the error rate of image recognition. Natural language processing (NLP) technology has made significant progress in 2013. Graves et al.[18] used recurrent neural network (RNN) for speech recognition work. In the same year, Mikdov et al.[19] proposed a neural network based language model word2vec for text analysis. Both techniques provide significantly better recognition results than traditional methods. The generative adversarial networks (GAN) technology born in 2014 is particularly concerned by the academic and industry[20], and the latest GAN algorithm has achieved a realistic effect in the field of image generation that is difficult to distinguish by the human eye. The deep reinforcement learning model Deep Q-Network (DQN) was published in *Nature* in 2015, marking a milestone in reinforcement learning and deep learning[21]. In 2016, AlphaGo, which combines deep neural networks, reinforcement learning, and Monte Carlo tree search, was developed by Google DeepMind and successfully defeated several Go champions. At the end of 2018, Google released a bidirectional language model, bidirectional encoder representations from transformers (BERT)[22], which opened the "Pandora's box" of deep learning in NLP application, and attracted great attention and wide application in the industry. It has become an important stage in the development of NLP technology. In 2020, Open-AI developed the pre-trained model GPT-3 with 175 billion parameters under the generative pre-trained transformer (GPT) system, which has become the strongest general language model in the NLP field, and showed close to human capabilities in translation, question and answer, text filling, and other application tasks[23]. In recent five, years, the data privacy security has gradually attracted global attention[24], and the "data silo" effect has become the "stumbling block" impeding the big data integration and AI development. In order to reconstruct industry data ecology, "federated learning"

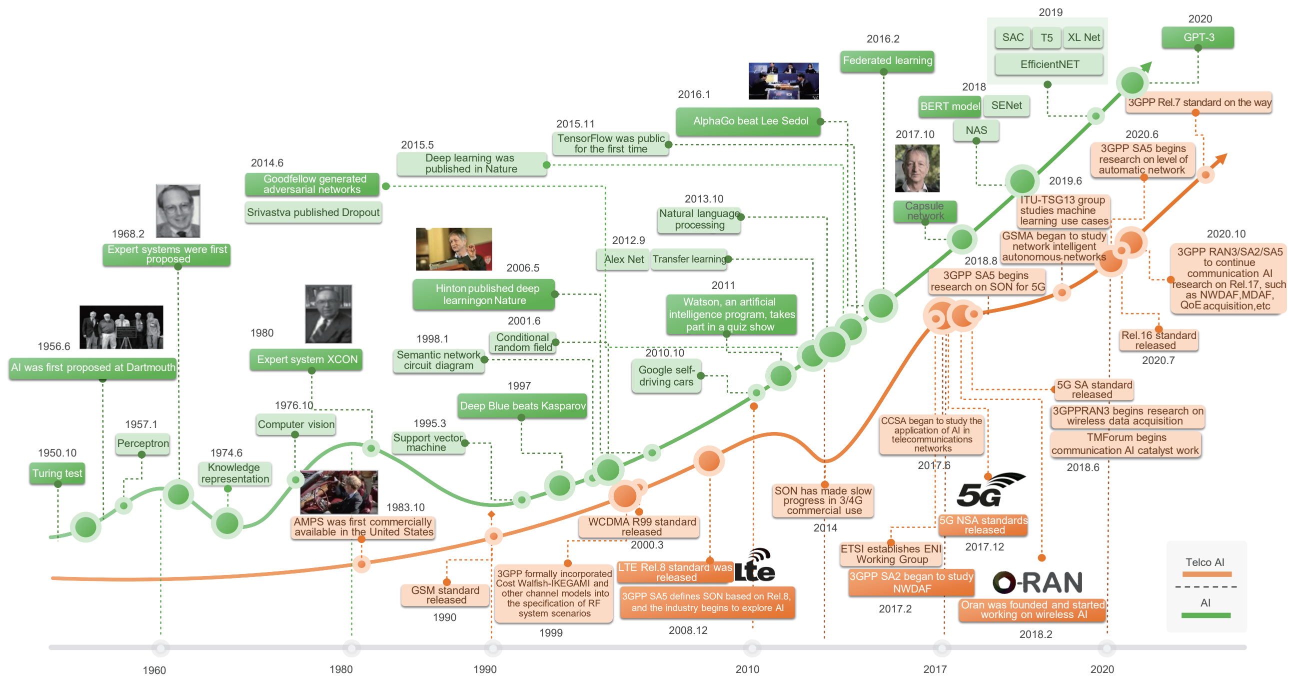


Fig. 1 Comparison of evolution path between communication technology and AI technology.

technology was first proposed by Google at the end of 2017[25], which broke the "data silo" deadlock with a distributed encryption machine learning method. In order to meet the demand of industrial data federation, WeBank[26] proposed an industrial-level federated learning framework, FATE, which realized a new paradigm of industrial federated learning in 2018. At the end of 2020, the industrial federated learning international standard IEEE P3652.1 was released, marking the official opening of the industry's "federated intelligence ecological alliance".

2.2 Historical development roadmap of telecommunications AI

Influenced by algorithms, computing power, requirements, etc., early mobile communication systems (such as AMPS, GSM, etc.) did not involve AI applications. However, the analytical methods based on data models and simulations have been applied in network planning and optimization. In 1968, Okumura[27] proposed the Okumura model and simulate the real wireless channel based on the real measured data, which can be regarded as the prototype of the early application of data science algorithms in mobile communication systems. In 1980, Hata[28] proposed the Hata model to optimize the Okumura model.

As shown in Fig. 1, 3GPP formally incorporated channel models such as COST Walfish-Ikegami into the specification of 3G radio-frequency (RF) system scenarios in 1999[29]. Later, with the development of wireless cellular technology, more wireless channel modeling and simulation algorithms emerged[30–32].

3GPP started defining self-organizing network function in 2008[33,34], then the communication community started exploring the application of various AI algorithms to SON. Initially, distributed optimization algorithms such as genetic algorithm, evolutionary algorithm, and multi-objective optimization algorithm are mainly used to optimize network coverage and capacity[35,36]. Machine learning has been widely accepted by the SON field as a key method for networks to achieve self-organization, self-configuration, self-optimization, and self-healing[37]. However, truly rapid development of telecommunications AI started in 2017.

In February 2017, 3GPP Service & System Aspects Working Group 2 (SA2) started researching intelligent network elements of 5G core network, network data analytics function (NWDAF)[38], for example, user equipment (UE) level mobile management such as paging enhancement and connection management enhancement based on UE mobility pattern prediction is studied; 5G quality of service (QoS) enhancement such as user QoS parameter configuration optimization is studied; network load optimization such as user plane function (UPF) selection based on network performance prediction is studied. In the same month, ETSI established the experiential networked intelligence (ENI) working group to study experiential-aware network management architecture, use cases, terminology, etc.[39] In June 2017, China Communications Standards Association (CCSA) launched a research project on the application of AI in telecommunication networks[40]. In February 2018, Open Radio Access Network Alliance (O-RAN Alliance) was established to develop the framework, use cases, processes and interface specifications of wireless AI[41]. In June 2018, 3GPP Radio Access Network Working Group 3 (RAN3) started to study the data collection mechanism on wireless side[42]. Telecommunication Management Forum (TMForum) also began the catalyst work related to AI. In August 2018, 3GPP SA5 started research on 5G SON related topics[43]. In October 2018, 3GPP SA5 began the research on AI and defined a new management plane function: management data analytic function (MDAF)[44]. In June 2019, ITU Telecommunication Standardization Sector Study Group 13 (ITU-T SG13) initiated research on machine learning use cases[45]. In the same month, Global System for Mobile Communications Association (GSMA) started work on a white paper on the intelligent autonomous network case[46]. In June 2020, 3GPP Service & System Aspect Working Group 5 (SA5) started the research topic of level of automatic network[47]. In the same month, China Mobile, together with AsiaInfo, officially introduce the federated learning concept in the 3GPP R17 standard for the first time, forming the first global international standard for federated learning in 5G field[48]. In July 2020, after the 3GPP R16 was officially frozen, 3GPP RAN3, SA2, and SA5 will continue to promote research on standardization topics such as NWDAF, MDAF, quality of experience (QoE) related to AI for the new R17 version.

3 Development of Telecommunications AI

The essence of communication is to transmit the information encapsulated in the signal from starting point to the destination through various communication technologies (such as mobile communication, satellite communication, fixed network communication, etc.). The quality of communication is measured by whether the information can be reproduced accurately and perfectly from the sending end to the receiving end. AI, relative to the natural intelligence exhibited by humans and animals, allows computer or machine simulate human thinking and cognition such as the capacity of "learning" and "problem solving", while perceiving environment and taking the corresponding actions to successfully achieve the presupposed objectives with the expectation of maximizing probability[49,50]. Communication and signal processing system are constructed based on sophisticated mathematical model. In contrast, deep learning in AI enables AI to absorb knowledge from data knowledge and make decisions without explicit mathematical modeling and analysis. If the mathematical assumptions are too elaborate and elegant, the communication system is easily deviated from reality in real-world applications. However, if AI or deep learning is applied to communication systems, its blackboxing learning process will easily make the construction of communication and information models lack physical meaning. Fortunately, a typical feature of communication systems is hierarchical autonomy and interconnection through standardized defined interfaces to form a complete system. For example, the transmitter and receiver of signal processing system can be decomposed into different processing units, responsible for their respective functions such as signal coding and decoding, channel coding and decoding, modulation, demodulation, noise removal, etc. The above scenario is very similar to the concept of micro-services in today's IT systems. Although such system architecture is not globally optimal, its advantage lies in the independent analysis and optimization of each subsystem to form an overall stable system. After more than 30 years of development, modern mobile communication systems have achieved excellent efficiency and performance, approaching the Shannon limit. Different from traditional hierarchical autonomous method, if AI and deep learning are used to consider the communication system as a whole model for analysis and optimization, it is possible to push the development of intelligent communication system to a new stage. This section will introduce the development and application of AI in various fields in the communication ecosystem, as well as the hierarchy defined by the working groups of international communication standards for the current development of AI in communication.

3.1 Development of telecommunications AI in the field of mobile network infrastructure

The development of AI in telecommunication network infrastructure is expounded from four aspects: wireless access network, core network, transport network, and terminal.

3.1.1 Wireless access network

The physical carrier of the wireless access network is base station. 5G base stations are divided into the central unit (CU) and distributed unit (DU), similar to traditional baseband unit (BBU), which are connected to the active antenna unit (AAU) through optical fiber. AAU contains traditional remote radio unit (RRU) and antenna function, namely, the active RF part is integrated with the passive antenna. For the CU, DU, and AAU in the wireless access network, AI is currently applied in the physical layer, MAC layer, and network layer[12].

In the physical and data link layers, the typical AI applications include using deep learning or reinforcement learning algorithms to evaluate and predict channel quality, detection of orthogonal frequency division multiplexing (OFDM) symbols at the receiving end, channel coding and decoding, dynamic spectrum random access, and other functions. The channel quality assessment uses some kind of algorithm such as the deep neural network (DNN) algorithm to analyze the limited pilot signal to help the massive multiple-input multiple-output (MIMO) system infer complete and accurate channel state information (CSI). The detection of OFDM symbols at the receiver usually relies on receiver evaluation using maximum likelihood estimation, but this method is very sensitive to CSI errors and the accuracy of the model itself. Therefore, Refs. [51, 52] tried the DNN algorithm and showed that the results surpassed the traditional MIMO symbol detection methods. In the 5G channel coding and decoding, low-density parity-check (LDPC) code is used for the data channel, and polar code is used for the control channel. Polar codes require multiple iterations to converge and reach optimized performance, while LDPC codes have higher decoding complexity under large blocks or noisy harshness. Thus various decoding algorithms based on Convolutional neural network (CNN), DNN, and reinforcement learning[53–56] present outstanding performance, quality, and small computational cost. The field of dynamic spectrum random access can also try the reinforcement learning-based random access and dynamic spectrum access (learning-based DSA strategy) in the future to ensure the dynamic spectrum access for large-scale terminals.

For the application layer of wireless access network, 3GPP defines the SON standard system[33,34,57,58], aiming to realize the self-configuration, self-optimization and self-healing of wireless network. Although 3GPP does not specify or suggest any statistical or data science algorithms, it suggests a series of intelligent SON application scenarios: network coverage and capacity optimization (CCO), energy saving management (ESM), remote control of electrical tilting antennas (RET), interference reduction (IR), automated configuration of physical cell identity (ACPCI), mobility robustness/handover optimization (MRO/MHO), mobility load balancing (MLB), random access channel (RACH) optimization, automatic neighbor relation (ANR), inter-cell interference coordination (ICIC), random access channel optimization (RACO), load-balancing optimization, self-healing functions, cell outage detection & compensation, minimization of drive-tests (MDT), etc. Since the first 3GPP SON standard in 2008, scientists in the industry[59–76] have tried a variety of data science algorithms, aiming to use AI methods to implement the above-mentioned multiple SON application scenarios. The overall development of SON has been relatively uneventful in the 12 years from the 3G to 5G era. The good wish of the communication standard makers is to inject automatic and intelligent genes into the mobile communication network architecture itself through SON, whether through distributed-SON (D-SON), centralized-SON (C-SON), or hybrid-SON (H-SON). Traditional communication equipment suppliers prefer SON to become a tightly coupled part of their own network infrastructure, so most of the SONs of traditional equipment suppliers support their own DE wireless network equipment.

Telecommunication operators prefer to deploy vendor-neutral and technology-neutral intelligent SON systems in their wireless access networks. Emerging SON startup companies are willing to stand together with communication operators to develop network-neutral SON technology, but they suffer from the closed and non-standard settings of functional interfaces and data interfaces by wireless equipment manufacturers, making it difficult to realize the 3GPP SON vision in the commercialization process. The relatively landmark event in the industry is the acquisition of Israel's Intucell by Cisco in February 2013 for $475 million, which was an SON-focused telecommunication software star company at that time. After 7 years of dismal development, Cisco sold to India's HCL in June 2020 for just 10% of the then-acquisition price, or $50 million. Amdocs, the world's top-ranked telecommunication software company, acquired US-based SON startup Celcite for $129 million in November 2013, but it did not achieve the expected commercial market share in the subsequent commercial SON process by the top two US mobile operators Verizon and AT&T. The American mobile communication operator has also gradually adopted self-developed mode to try in the field of intelligent wireless access network in the past 5 years. For example, Verizon's SON system jointly deployed with Cisco and Ericsson in 2015 was gradually replaced by its own V-SON system.

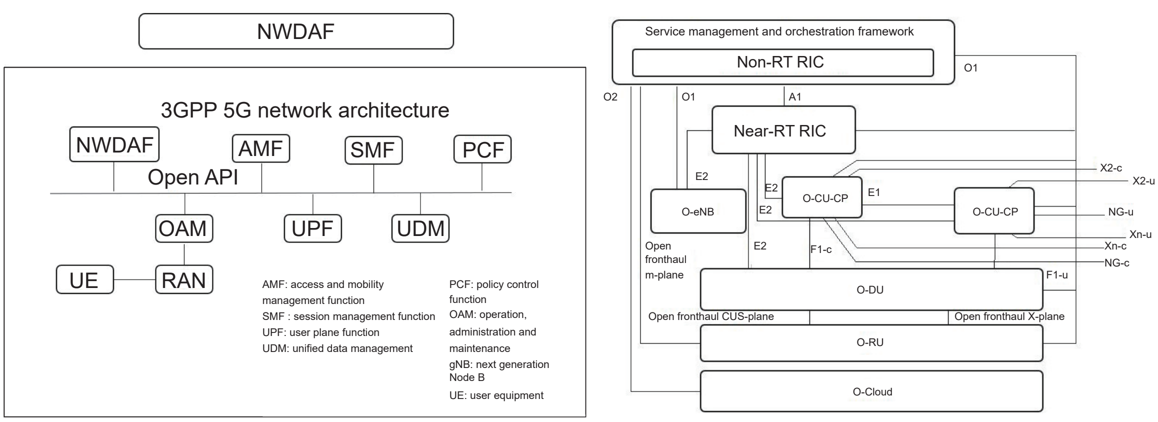
3.1.2 Core network

The development of AI in the field of mobile communication core networks has made significant development in the 5G era. 3GPP SA2 defined NWDAF in February 2017, which is the first time that mobile communication has been defined and standardized in the core network architecture since 1G to 5G, and requires the deployment of network AI network elements. The NWDAF framework is shown in Fig. 2a, the NWDAF aims to use AI algorithms to integrate with communication technology protocols to intelligently manage, optimize, and improve network quality and experience for mobility management, QoS, and other network elements (such as UPF) of the 5G core network. Currently, Chinese and American operators are conducting functional testing of NWDAF for commercial use in 5G standalone (SA).

Compared to 3GPP, O-RAN is another emerging technology track. Since 2018, O-RAN Union has formulated AI-enabled RAN intelligent controller (RIC), and tightly coupled with management & orchestration (MANO) function of the core network. Therefore, we expound the RIC in the core network part. RIC framework is shown in Fig. 2b. The RIC is divided into non-real time RIC (Non-RT RIC) and near-real time RIC (Near-RT RIC). The Near-RT RIC is sunk on the radio access network side and connected to the CU/DU through the E2 interface. The functions of near-real time RIC include wireless resource management, mobility management, wireless connection management, switching management, and wireless QoS management using AI capabilities. Non-RT RIC is defined in the core network MANO system, and is connected to the Near-RT RIC through the AI interface, and its main functions are the AI-based service and policy management, high-level service process optimization, helping Near-RT RIC offline training of AI models, etc. RIC of O-RAN is still in the process of standard refinement and early trials. Compared with 3GPP NWDAF, O-RAN RIC is still a long way from mature commercial.

3.1.3 Transport network

The transport network is the foundation of the communication network, which is responsible for physically connecting the various network nodes and transferring the data to the destination. Because optical communication has the characteristics of large bandwidth, stability, and low loss, the current mainstream transmission technology is to communicate through a carrier such as an optical network. Optical transport network development has gone through the development and innovation of plesiochronous digital hierarchy (PDH)[77], synchronous digital hierarchy (SDH)[78], multi-service transport platform (MSTP)[79], wavelength division mulplexing (WDM)[80], automatically switched optical network (ASON)[81], and optical transport network (OTN)[82] technology. In order to achieve flexible control of network traffic and better support the needs of service transmission, the industry has begun to explore the introduction of software defined network (SDN) into optical networks to realize the software defined optical networking (SDON)[83] in recent years. SDON inherits ASON's characteristic of dynamically recovering from service interruption, but at the same time, it is also committed to guaranteeing network capacity and service reliability. In addition, due to the need to reduce operating costs and provide the automation and intelligence level of optical network networking, optical networks need to be combined with technologies such as big data, AI, and cloud-network convergence to realize intelligent management of optical networks. This also introduces the concept of cognitive optical network (CON) and the industry exploration[84]. According to the objectives of the EU-funded cognitive heterogeneous reconfigurable optical network (CHRON) project[85,86], the core of the CON is the cognition and decision system (CDS), which is responsible for managing transmission requirements and network



(a) NWDAF framework (b) RIC framework

Fig. 2 NWDAF and RIC frameworks.

events. The control and management system (CMS) is responsible for controlling and disseminating relevant commands, as shown in Fig. 3. Currently, there are some research results about the combination of SDON/CON and AI such as predicting faults, shortening recovery time improving the signal-to-noise ratio of light[87], etc.

In addition, the evolution of Internet protocol version 4 (IPv4) to the Internet protocol version 6 (IPv6) focuses on solving many problems such as the address space and QoS guarantee of network transmission, etc. In order to meet the needs of 5G scenarios, building an IPv6-based intelligent IP network is also the development trend of bearer network. How to achieve flexible network routing, ensure network slicing service level agreement (SLA) of bearer network and deterministic network transmission requires the use of AI technology for guarantee. For example, the IP network intention can be identified and judged based on AI technology to guarantee the network experience in a targeted manner, but in general, the current combination of IPv6 network and AI is still in the initial exploration stage. The industry hopes to use AI technology to monitor the operation state of the whole network, timely detect network problems and risks and intelligently identify the network anomalies, locate the root cause of faults and generate relevant optimal strategies for problems found. To better realize the intelligent IP network, it is also necessary to introduce technologies such as IPv6, segment routing IPv6 (SRv6), and flow-following detection[88,89] to provide network-aware and flexible routing configurations, and upgrade the IPv6 technology to IPv6+. At present, relevant research in the industry is still in the exploratory stage.

Currently, the resource allocation and storage computing resources are relatively independent at the cloud network side. For example, complex AI applications are performed on cloud servers, and simple and lightweight AI applications are performed on terminals. With the development of SDN, IPv6, IPv6+, and other technologies, the industry is committed to realizing new architectures such as the integration of computing power networks and IP networks, and cloud-network integration. In this process, there are still many technical problems such as how to achieve the optimal routing, how to distribute computing power optimally, and how to guarantee the service quality of computing power, etc., which requires the use of AI and other technologies to overcome the difficulties. Now, the relevant research is still in the initial stage.

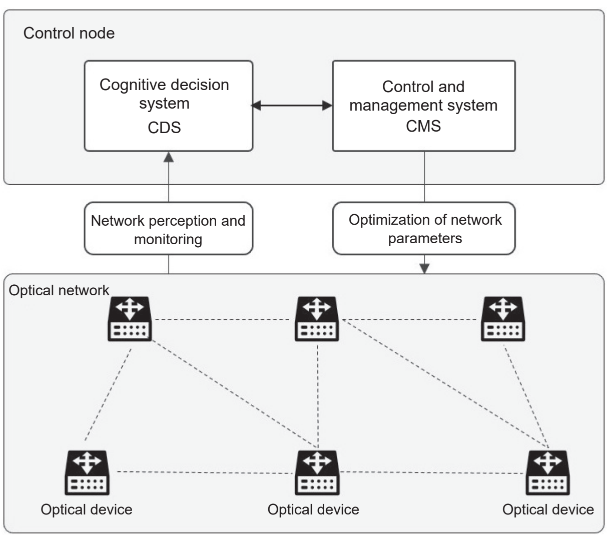


Fig. 3 System structure of cognitive optical network.

3.1.4 Terminal

Terminal-based AI includes the intelligence of terminals and chips. The terminal operation system itself and the Apps at the application layer have already developed some intelligent application, while the empowerment of network infrastructure by terminal-based AI, which we focuses on, is still in the early development stage at present. A typical application is that the performance data collected by the terminal chip is reported to the SON system or the operation supporting system (OSS), and the network AI analysis engine of these two components is used to intelligently optimize the wireless network. In the 3GPP standard, it is reflected by the minimization of drive test (MDT) in the 3GPP SON.

Overall, the development of AI in the field of communication network infrastructure was relatively uneventful in the 3G/4G era, while it accelerated in the 5G era. With the gradual maturity of NWDAF of 3GPP, RIC of O-RAN and RAN data analytics function (RAN-DAF) of 5G Mobile Network Architecture (5G-MoNArch) projects group[90] and accelerated commercial progress, AI will be further deeply integrate into network architecture of 5G and B5G, and will exist and evolve in the form of independent network element and network function entities for a long time. Meanwhile, it can also be seen that the AI mathematical model and mobile communication field knowledge are still relatively independent, that is, many of the mathematical model results still lack the physically meaningful explanation at the communication level, and the two need to do further deeply integrated to enhance the interpretability of AI in communication physical network application. The traditional communication equipment manufacturer also needs to open data interface black box, and help operators to build a neutral (vendor agnostic) intelligent network infrastructure.

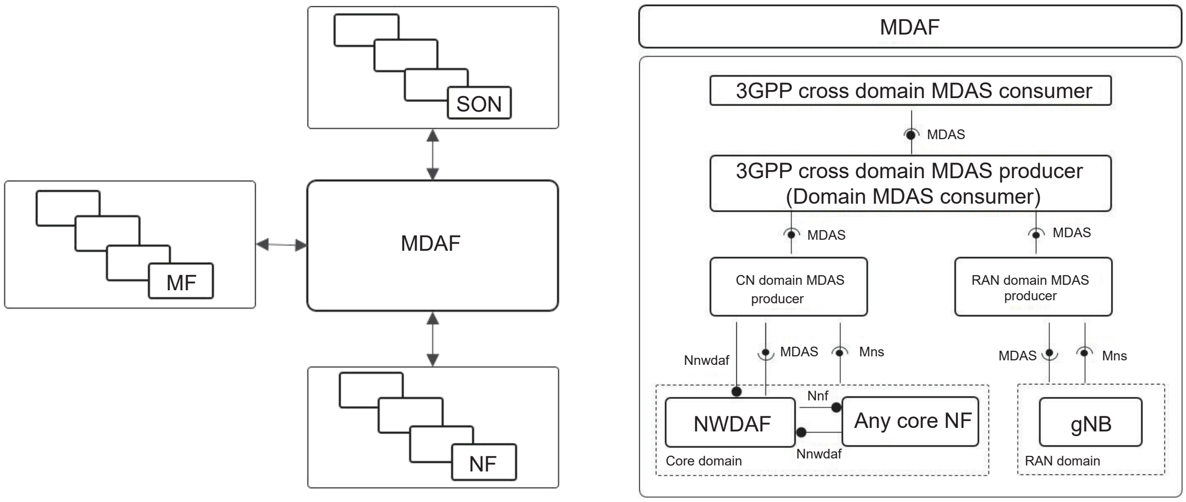
3.2 Development of telecommunications AI in network management field

The development of AI in network management field is described in three parts: MDAF, ENI engine, and network OSS.

3.2.1 MDAF

The MDAF and the service object defined by 3GPP R16 are shown in Fig. 4a. 3GPP SA5 starts defining intelligent function of network management in 3GPP R16, such as the management data analysis function (MDAF), which helps the management system to set reasonable network topology parameters for network configuration and guarantee the quality of service by conducting data analysis. After the network is properly configured according to the analysis results provided by MDAF, the control plane and user plane can conduct further parameter adjustments to improve user experience. For the operation administration and maintenance (OAM) system, the key step is business requirement analysis and what information should be provided to MDAF. For example, communication service management function (CSMF) of network slice translates the customers' SLA into communication service demands, and uses the analysis capability of MDAF to judge whether the requirement matches the existing slicing instance, and guarantees slicing SLA by selecting the optimal slicing instance[44].

MDAF can also empower SON on management plane, as shown in Fig. 4a. MDAF utilizes the collected management plane and network data to perform relevant analysis and implement various SON functions described in Section 3.1. Nevertheless, MDAF's standards (such as the definition of interface, collected



(a) MDAF defined in 3GPP R16 and the service object (b) MDAF architecture defined by 3GPP R17

Fig. 4 MDAF defined in 3GPP R16 and the service object[87] and MDAF architecture defined by 3GPP R17[87].

data information and process, etc.) in 3GPP R16 are not well defined and it is relatively difficult to apply and deploy in current 5G network. In response to the problems of 3GPP R16 and the new scenario, 3GPP starts enhancing MDAF in 3GPP R17, defining and improving scenarios such as the coverage enhancement, resource optimization, fault detection, mobile management, energy-saving, paging performance management, and SON cooperation, etc., in addition to improving the function in 3GPP R16[91]. The relationship between MDAF and service object defined in 3GPP R17 is shown in Fig. 4b. The service architecture of MDAF in 3GPP R17 is more complex and how it collaborates with NWDAF has not been finalized yet, and its technical standard in 3GPP R17 is in progress. MDAF has not yet been commercially deployed in the 5G network management system of Chinese and American operators.

3.2.2 ETSI ENI

ETSI defined the ENI system in 2017 as an independent AI engine to provide intelligent service for applications such as network operation, network assurance, equipment management, service orchestration and management, etc.[39] ENI's functional framework is shown in Fig. 5. ENI contains context-aware, knowledge management, cognitive processing, situation-aware, model-driven, policy management, and other AI-related knowledge management, model management, and policy management modules. The raw data is cleaned and feature processed through input processing and normalization, and then processing through the internal AI module, it provides policy or instruction to OSS, business support system (BSS), users, system applications, orchestrators, infrastructure, and other service objects. The module formed by deregularization and output translates the policies or instructions generated by ENI, and outputs the language that the service object can understand.

Currently, ETSI ENI is still evolving the function of ENI, and defines more advanced applications such as the energy saving based on intent based network[93], data mechanism[94], matching of ENI and operator system[95], etc. The relevant work is still ongoing. For the function and application scenarios defined by ENI, the domestic and foreign operators have tried related pilot projects, and achieved good results in slice management, user experience optimization, wireless energy optimization, etc.[96] For example, in the pilot report shown in reference[97], companies such as UC3M,

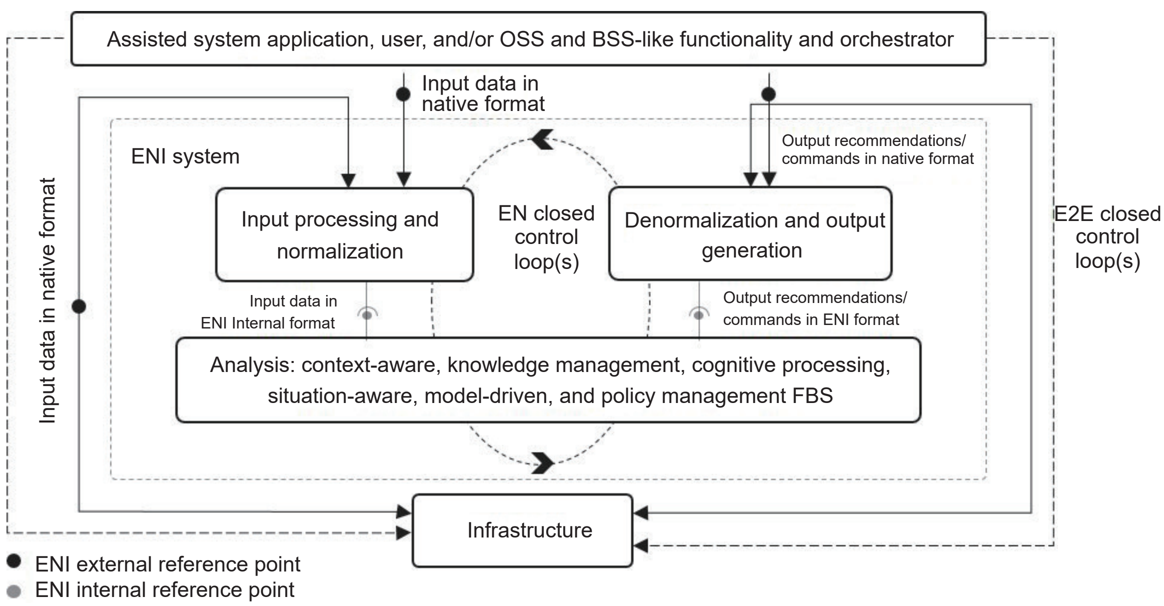


Fig. 5 ENI functional architecture[92].

Samsung, and Telecommunication Italia have cooperated to implement elastic management and control of slice resources through ENI, which can well control the performance of end-to-end delay, service creation time, and system capacity. In the report presented in Ref. [98], China Telecom, AsiaInfo, Beijing University of Posts and Telecommunications have cooperated, and they use ENI for intent-based user experience optimization and improve user experience management. ENI is not yet deployed as a separate AI system or network element in 5G networks or network management systems, but many of its defined functions in the form of decoupling are starting to have applications within the network management systems of global operators.

3.2.3 Network OSS

The network OSS is often defined as a kind of software function that enables communication operators to manage its network and application. An OSS system usually has at least five major functions: network management system, service delivery, service fulfillment, which including network inventory and network activation and provisioning, service assurance, and customer care[99]. The early development of AI in the OSS field was also very slow. In the 1970s, the vast majority of OSS work was still performed manually and with human intervention; in the 1980s, with the rise of Unix system and the C language, Bell System started developing OSS systems, including some of the most famous early OSS systems such as automatic message accounting tele-processing system (AMATPS), centralized service order bureau system (CSOBS), engineering and administrative data acquisition system (EADAS), switching control center system (SCCS), service evaluation system (SES), etc. AI did not have any application in those early systems. In the 1990s, ITU-Telecommunication Standardization Sector (ITU-T) defined a new four-layer architecture of OSS in its telecommunications management network (TMN) model: business management level (BML), service management level (SML), network management level (NML), and element management level (EML). Later, fault, configuration, accounting, performance, security (FCAPS) was introduced as a new model to manage these four levels. In the commercial process of TMN, AI was still rarely seen. After 2000, with the development of new generation operations systems and software (NGOSS) project and enhanced telecommunication operation map (eTOM) architecture, OSS framework system of communication operator has relatively standardized guidance standards. The application of AI also has a relatively stable framework and carrier, especially for the fault management and performance management in FCAPS, AI has started to be tried and applied in the areas of fault diagnosis, root cause analysis, performance prediction, etc.[92]

OSS attributes defined by TMForum is shown in Fig. 6. In 2019, TMForum defined in the Future OSS research report that future OSS is “data-driven” and must rely on AI, machine learning, automation, micro-services, and business optimization, and must has the characteristics of agility, automation, initiative, predictability, and programmability[100]. Among 10 most important factors that define the future of OSS, 4 factors are closely related to AI: execution and assurance of automated closed-loop service process, optimization of automated closed-loop network, AI-driven customer interaction, and AI-driven network optimization. Therefore, mainstream communication operators also gradually embedding AI platforms or functional modules in the OSS for 5G evolution in anticipation of the intelligent evolution of OSS. As shown in Fig. 7, AT&T’s enhanced control, orchestration, management & policy (ECOMP)[101] system defines the AI-based analytic application design studio (AADS) function in the design state. In the running execution state, ECOMP also defines the function of data collection, analytics and events (DCAE), which provides the AI-based real-time FCAPS functions, manages and orchestrates services, networks, and resources through intelligent analysis, and realizes automated closed loops[102]. Verizon also take AI as a must-have function of BSS/OSS, especially in the user network and business experience has been deeply applied[103].

China’s operators insert a new platform or component between data middle-office and OSS core functional modules in the 5G OSS-oriented system construction, which is named as network AI middle-office or intelligent middle-office to undertake the network AI function. In Fig. 8a, we abstract 5G OSS network middle-office system of China’s three major telecommunication operators into a technology-neutral 5G OSS network middle-office architecture, in which the data middle-office is mainly responsible for network-side data collection, data storage, data governance, data sharing, and other functions. In order to meet the intelligent requirements of 5G network, business, and service management, the network AI middle-office uses the network big data in the data middle-office as the main resource, and continuously builds, infers, publishes, and precipitates network AI algorithm models around the planning, construction, optimization, operation, and other scenarios in the network life cycle. It provides network AI functions including anomaly detection, capacity prediction, network optimization, root cause analysis, alarm prediction, fault self-healing, service orchestration, and perception optimization for 5G networks, and comprehensively improves the automation and intelligence capabilities of 5G networks.

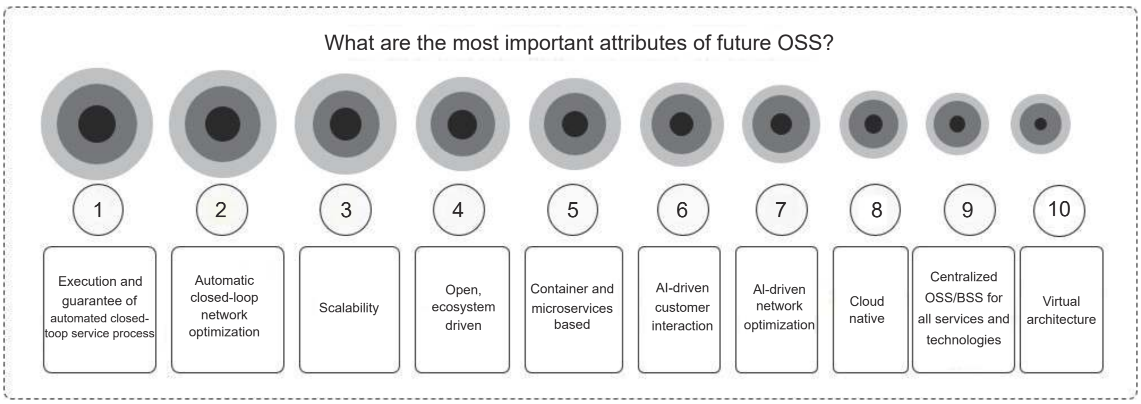


Fig. 6 OSS attributes defined by TMForum.

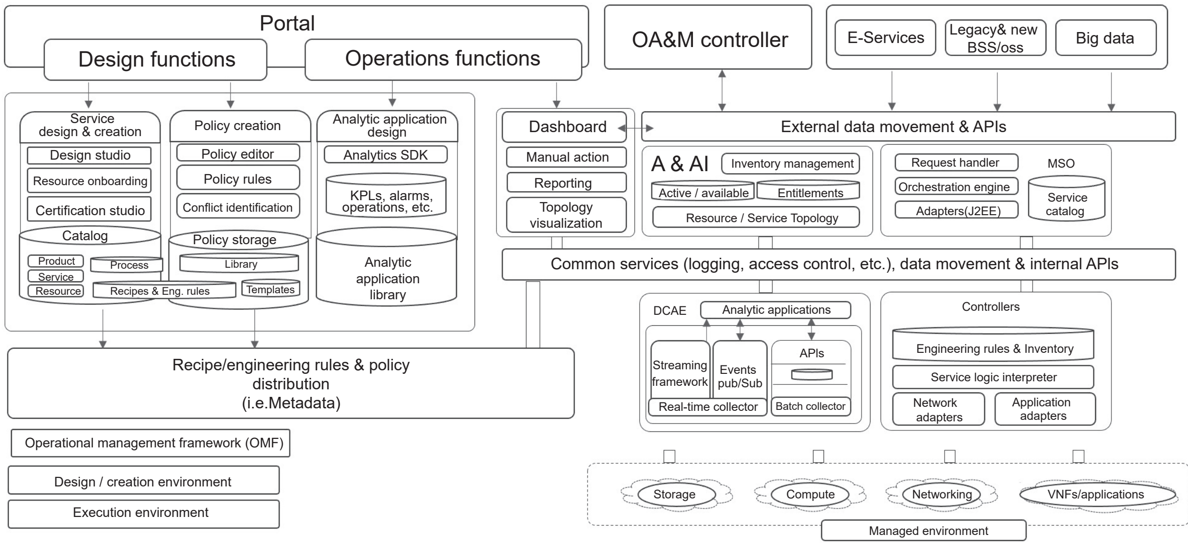
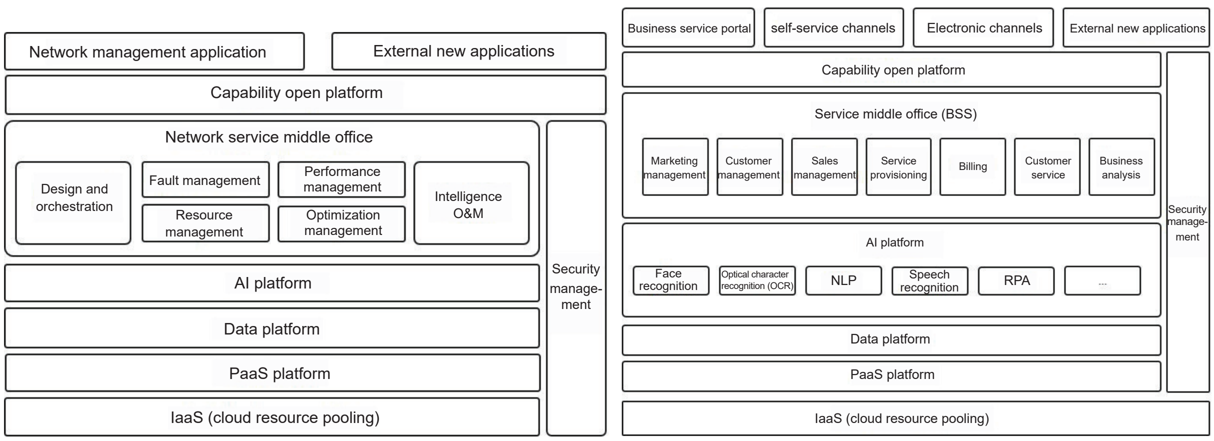


Fig. 7 ECOMP defined by AT&T.



(a) 5G network operation support system   (b) 5G business support system

Fig. 8 5G network operation support system and 5G business support system.

3.3 Development of telecommunications AI in the field of the telecommunications business

The telecommunication business supporting system provides the capabilities for customer operations and services. As shown in Fig. 8b, BSS functions in TMForum's framework system are mainly oriented to the customer's market and sales, products, customers, services, resources, suppliers and partners and other fields[103]. The core production system of BSS includes customer relationship management (CRM), billing system, business intelligence (BI) system, call center (CC) system, etc. The BSSs of mainstream operators around the world have been perfected, and most of them have completed centralized and platform-based construction. Chinese operators are leading the technological evolution based on smart middle-office in the BSS field, that is, through the ability to operate and coordinate business middle-office, data middle-office, technology middle-office, AI middle-office and other middle-office systems to complete IT services and interactions for end users and partners, where the AI middle-office is based on AI algorithms, and through the encapsulation of scene-based service capabilities, it adds intelligence and empowerment to the business process. Up to now, AI technology has been well applied in various business fields and corresponding scenarios such as marketing, sales, customer service, and billing in the BSS domain through the AI middle-office system.

3.3.1 Marketing

The typical AI application in telecommunication marketing is customer marketing intelligent recommendation and operational decision assistance. Through various AI recommendation algorithm models and expert experience rules, a targeted recommendation strategy model is formed, and the best matching strategy is output according to customer characteristics. On this basis, with the help of AI decision-making related algorithms, a comprehensive decision-making model is constructed by combining factors such as product matching degree, value degree, and company benefit, so as to generate the best operational decision and help enterprises improve efficiency. Applicable scenarios include recommendation of popular products, recommendation of relevant products, recommendation of personalized package, recommendation of contract, recommendation of digital content, etc.

3.3.2 Customer sales

The typical AI application in this field is the use of face recognition, OCR and other technologies to support scenarios such as identity authentication and auditing of business acceptance by customers in business halls, sales agreement signature authentication, and real person business processing confirmation. In the sales process of government and enterprise customers, OCR and image recognition technologies are used to support the automatic identification of enterprise information in the business record link, the identification of government and enterprise business seals to achieve pre-authentication and audit, the automatic drafting of sales contracts and other scenarios to realize the intelligent and automated business sales, and improve the work efficiency of account managers.

3.3.3 Customer service

The typical AI application in this field is based on technologies such as speech recognition, intent recognition, multi-modal question-answer matching, speech synthesis, semantic processing, user sentiment analysis, tag multi-classification prediction, OCR, and other technologies to achieve voice interaction between customers and intelligent robots, and real-time monitoring of customer sentiment. The application also includes predicting customer demands and effectively assigning service agents, monitoring customer problems in real-time, performing automatic classification and identification, automatically retrieving knowledge bases, and assisting agents to reply to questions. Applicable scenarios include: identity authentication based on customer voiceprints, prediction of potential complaints, voice quality inspection during the customer service process, intelligent quantitative scoring, intelligent dispatch based on work order text information, automatic generation of knowledge from knowledge base, intelligent customer service scheduling, etc.

3.3.4 Billing

The typical application of AI in this field is to use AI for IT operations (AIOps) to support the grayscale release of billing system application process upgrade and configuration modification online, so as to realize fault discovery, fault diagnosis, and fault detection, fault self-healing, fault prevention of billing system; combined with the robotic process automation (RPA) capability, it supports the automation of daily billing in the billing system. In addition, based on multi-dimensional pricing factors (bandwidth, latency, reliability, accuracy, number of connections, capacity, number of network function instances, etc.) and customer data, AI algorithms can be used to achieve intelligent pricing and determine the optimal price to help business-to-customer (B2C) enterprise maximize their benefits.

3.4 Development of telecommunications AI in cross-domain integration intelligence

No telecommunication service can be operated or supported in the single system of OSS or BSS. The integrated telecommunication business process, the evolving technology middle-office system architecture features, and the integrated analysis and operation of business and network data are the three main factors that drive the deep integration of BSS and OSS. The development of telecommunication AI in cross-domain integration intelligence has also given rise to a variety of application scenarios and cases.

3.4.1 Customer experience management (CEM)

In the organizational structure of telecommunications operators, the network operation support and business operation support are usually relatively independent fields, and their corresponding production systems, OSS and BSS, also operate and evolve relatively independently. The network field focuses on various key performance indicators (KPIs) of network and network elements, while the business field is responsible for developing new customers and new services and serving existing customers for the market. Communication operators usually adopt the QoS system formulated by ITU and International Organization for Standardization (ISO) for their SLA signed with their customers. Traditional QoS is technology-driven and defines service quality by network KPI, and it cannot truly reflect the experience and feelings of users in using network services. Therefore, a “digital gap" is formed between the network performance quality based on the QoS system in the network field and the customer satisfaction and perceived experience that are concerned in the business operation field[104], as shown in Fig. 9. As a new field of cross-domain integration of network and business, CEM uses AI technology to evolve the operator’s QoS system to a user-centered quality of experience (QoE) system, realizing the transition from network KPI-centered to customer experience-centered network business service[105].

The core of CEM is to use AI combined with psychology to establish a set of algorithmic model system that can accurately reflect the customer’s perceived experience of telecommunication network and service usage, which we call telecommunication psychology algorithm. It quantitatively maps the QoS of the network system to the QoE system of user experience, thereby bridging the gap between network quality and real user experience. Currently, there are two common methods for evaluating user experience perception: net promoter score (NPS) and emotional connection score (ECS)[105,106], as shown in Fig. 10. NPS is used to measure the likelihood of customers recommending a certain company, product, or business[106,107]. It builds a scoring index system based on customer feedback, and surveys users’ satisfaction with a company, a product or service through telephone or questionnaire interviews, quantified on a scale of 0 to 10[106]. NPS is relatively a passive, static quantitative evaluation mechanism based on long-term customer impressions. Now, about 7% global telecommunication operators use the NPS to measure the customer satisfaction[108].

With the advent of the emotional connectivity score (ECS) model that incorporates psychology into machine learning, the traditional NPS indicator system has become obsolete. A large number of studies have shown that customers’ ECS is the most closely measured system of real experience quality[105]. Unlike NPS,

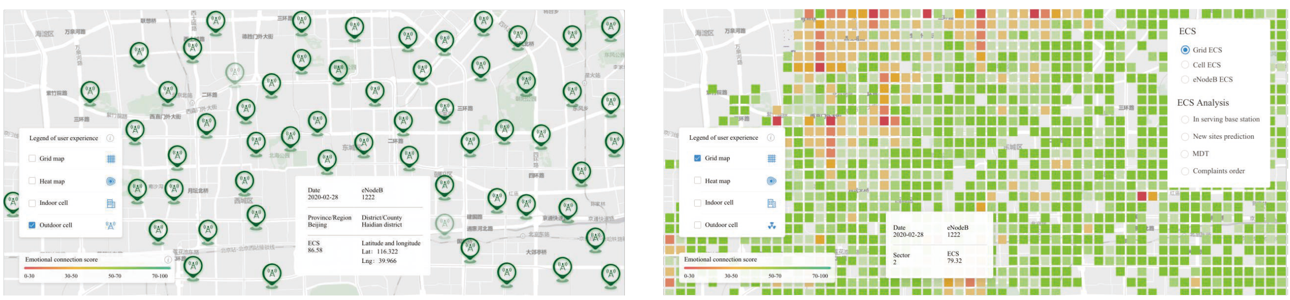


Fig. 9 Gap between performance and experience.

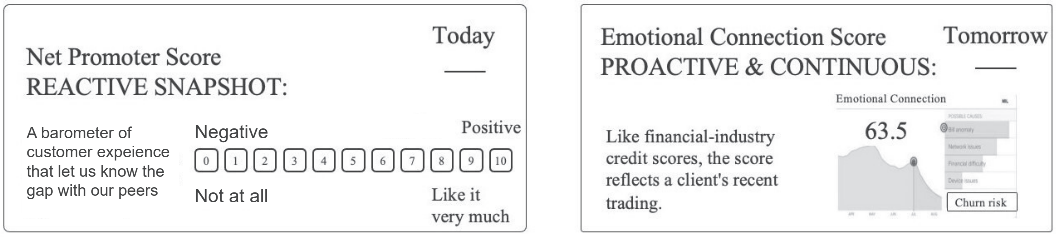


Fig. 10 Indicator system comparison: Net promoter score (NPS) and emotional connection score (ECS).

ECS is active and continuous. As shown in Fig. 11, by aggregating and accessing multi-faceted data across networks and business areas, the model uses ECS telecommunication psychology algorithms to continuously learn quantifiable customer experience and the connection between the various indicators, and map the event results into ECS scores. In this way, operators can discover the root cause of customer experience problems in time, guide the improvement of network and service quality, and achieve the goal of improving user experience. AsiaInfo proposes a set of telecommunication psychology experience perception algorithm and an index set that can quantify user perception experience[106], and conducts massive data machine learning on user-level subjective data (e.g., NPS research, customer complaints, active dialing tests, etc.) and objective data (e.g., quality indicators of voice calls, Internet access, high defintion (HD) video services, virtual reality (VR) services, etc.) in the communication field, compares and verifies the difference of user data among various regions, and optimizes ECS parameters, then combines user-level communication, consumption, service, and other behaviors to conduct portrait analysis, and finally generates an ECS telecommunication psychological model, which is used to instantly evaluate the instantaneous experience quality of any business at any moment, anywhere, and any kind of business in the customer journey.

The application of telecommunication AI in the customer's life cycle journey in CEM, as shown in Fig. 12. Telecommunication AI takes the improvement of user experience as the convergence goal, evaluates the user's instantaneous service experience quality, and quickly locates and diagnoses the problem of experience degradation, runs through the user personalization strategy of active perception and care, so as to gain insight into the experience indicators of each user in the network journey, and provide personalized services to users through the network and business system to realize the 5G network personalization (NP).

3.4.2 Policy control function+ (PCF+)

Network policy control is mainly based on rule definition, and is implemented by the policy and charging rules function (PCRF) in 4G LTE[109]. 3GPP introduces PCRF network element from R7 (Release 7) to control quality of service for the user and service QoS, and provides differentiated service for users. Besides, it can provide users with service stream bearing resources guarantee and stream billing strategy, realize more refined service control and billing methods based on service and user classification, so as to utilize network resources reasonably. PCRF contains policy control decision and the function based on stream billing control, providing network control function of data stream detection, gate control and stream billing, which can generate control rules by multiple dimensions such as business, user, location, accumulative usage, access type, time as triggering conditions. AI has not been applied in PCRF, and the policy rules of PCRF are mainly based on rule configuration. The PCRF architecture is shown in Fig. 13a, which implements policy and charging functions by interacting the Gx reference point with network elements such as the policy and charging enforcement function (PCEF).

3GPP only defines the PCF for network side, but still needs to be combined with the integration of the BSS & OSS (BO) domain to enhance the function. The reason is that with the enrichment of 5G services and the increasing demand for differentiated 5G services, operators not only have higher requirements for the use of network resources and network control, but also need to have more detailed analysis capabilities combined with business strategies. This analysis capability often requires mastering user attribute characteristics or event information in the business domain. This requires that the scope of PCF's role be extended to collect data not only in the OSS domain but also in the BSS

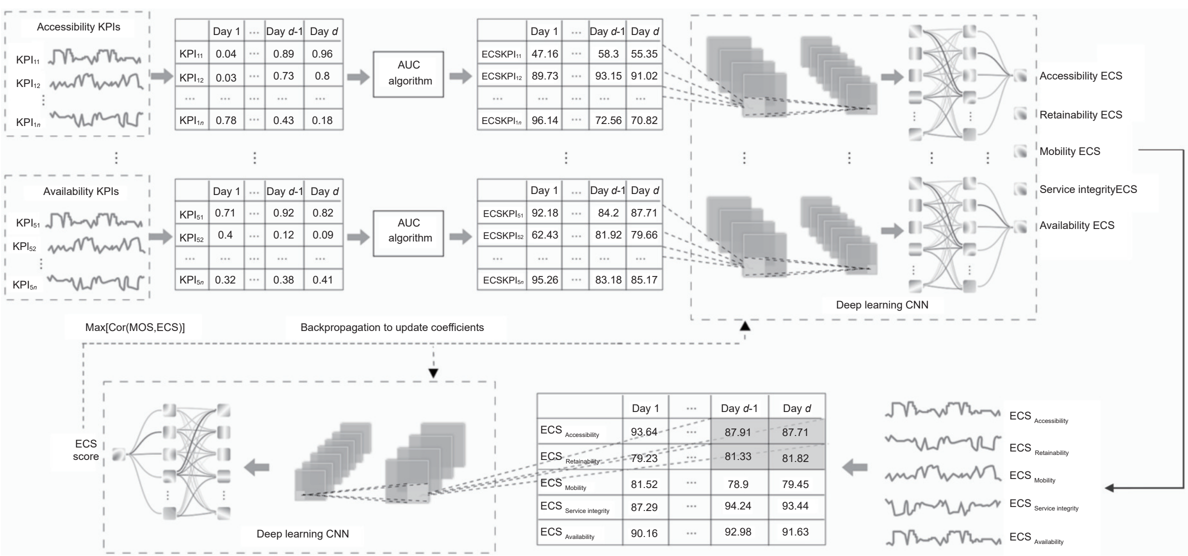


Fig. 11 ECS telecommunication psychology model.

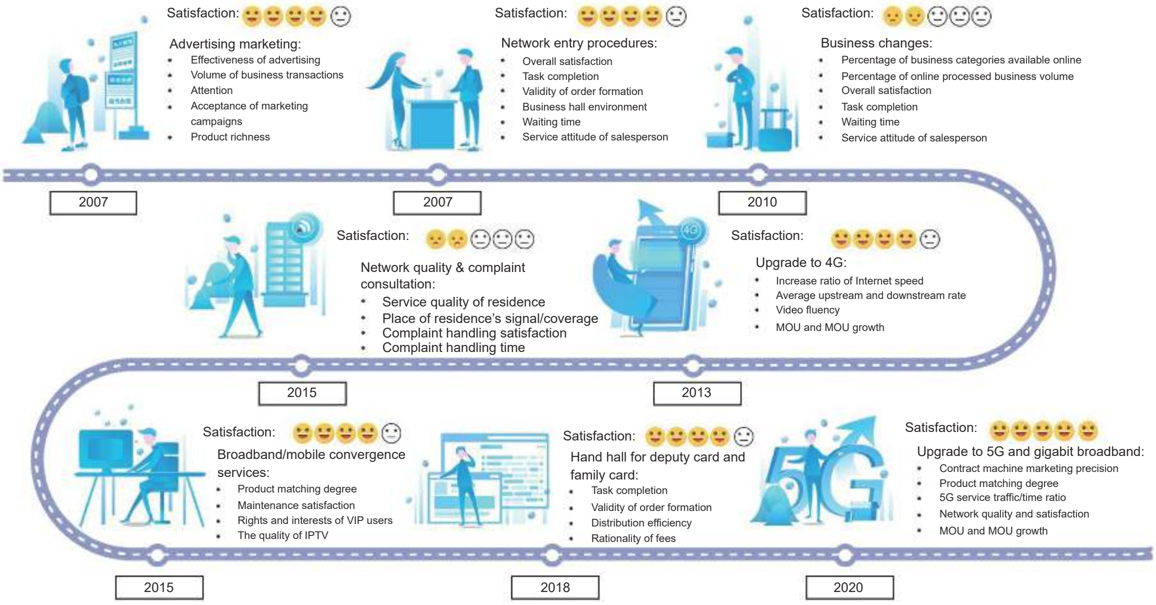
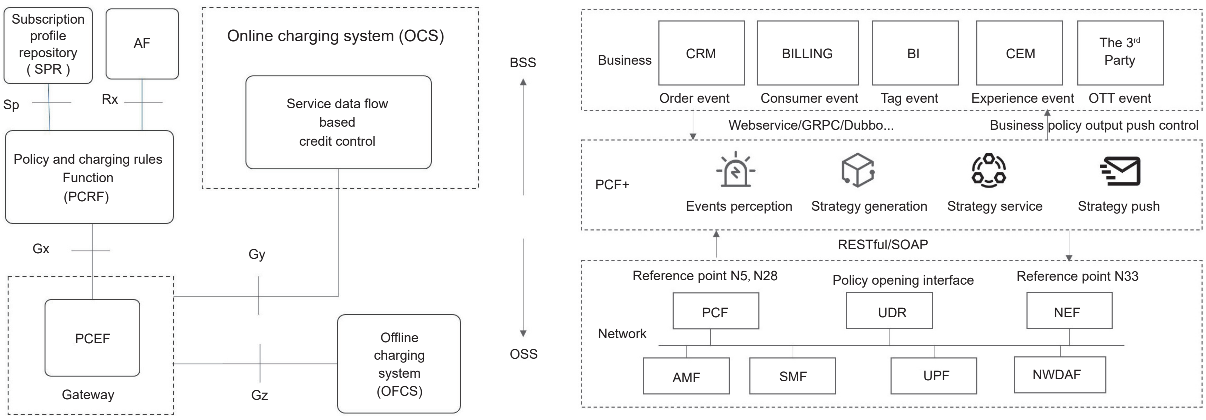


Fig. 12 Full life cycle view about customer experience.



(a) 3GPP PCRF framework (b) PCF+ based on BO domain integration

Fig. 13 3GPP PCRF framework and PCF+ based on BO domain integration.

domain, and the application objects should also be extended from the network domain to the business domain. Based on the above consideration, PCF needs to evolve to PCF+ to provide new business modes, business scenario or business model, as shown in Fig. 13b.

In addition, due to the introduction of AI/big data capabilities in 5G core (5GC), policy management and control have become more intelligent, rather than the previous PCRF-based expert system or rule configuration function. For example, through the integration of NWDAF and PCF/PCF+, intelligent slice experience management, intelligent SLA guarantee, and other functions can be easily realized. In addition, PCF+ can use the fusion data of the network domain/service domain (obtains the cell congestion status from the network domain, and obtains the user's level and package usage from the BSS domain), and dynamically adjusts the user's package settings or recommends the optimal package by combining AI technology to ensure user QoS and improve user satisfaction.

3.5 Development case of telecommunications AI in private network

5G provides vertical industries with both private and public network. Telecommunications AI offers series of customized service and security guarantee in 5G private network (PN). According to the sharing model of network resources, the 5G PN deployment is divided into 3 types: virtual PN, hybrid PN, and independent PN. As shown in Fig. 14, the virtual private network is based on the current 5G public network, realizes the service bearing of professional users with the slicing method, and has the following features: virtual private network and public network share the UPF network element; combined with customer needs, the private network can be connected to the enterprise through the existing virtual.

The hybrid private network is based on the core control plane and wireless access part of the 5G public network, and carries private network services, with the following features: it supports user plane data privatization, independent UPF and multi-access edge computing (MEC), and the equipment is built in the park; user access is authenticated by slicing identification.

The independent private network is completely isolated from the operator's 5G public network, and is characterized by: independent networking is achieved through the simplification of the core network and the combination of UPF and wireless base stations; private base stations ensure full coverage, and UPF/MEC ensures private network data sinking and network independence.

Telecommunication AI can be applied in any of the above 3

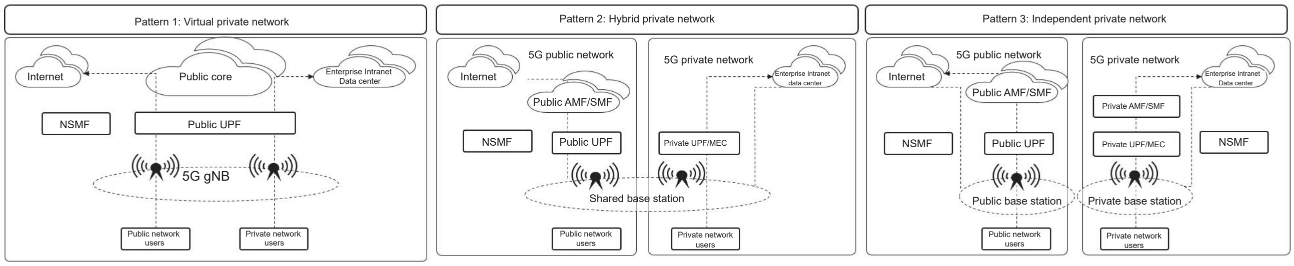


Fig. 14 Three kinds of models of 5G private network.

modes of private network mode. For example, it can be used for network slicing SLA guarantee in virtual private network to optimize communication transmission performance, quality, and resources. For independent private networks, the AI experience perception algorithm can conduct real-time or quasi-real-time evaluation of users' perceived experience, provide accurate QoS combinations to ensure service quality, and achieve differentiated intelligent operation services. In the private network, AI technologies such as federated learning and transfer learning can also be used to complete the cloud-based management and continuous learning optimization of the 5G slice abnormality diagnosis model. In addition, AI technology can also perform real-time assessment of wireless network performance in the private network[37]. It is used to improve the video quality of the application layer or accelerate the game by interacting with vertical industry application platforms and adaptively adjusting the parameter settings of the application layer.

3.6 Classification definition of telecommunications AI development by ISO

In view of the rapid development of AI technology since the commercialization of 5G, the Telecommunication International Standards Organization has also begun to initially grade the development maturity of Telecommunication AI. Intelligent autonomy cannot be achieved overnight and needs to be developed gradually. A complete intelligent autonomous network is the ultimate goal, but it also needs to start with automated repetitive operation methods. First, the automation of network operation is initially realized; second, the network environment and status are actively sensed, and machine learning is used to make continuous optimization decisions; then, in the process of developing from network perception to cognition, user intent is recognized, and a closed-loop cognitive learning network is constructed; finally, a closed-loop autonomous network body from perception to cognition and then to prediction is realized, and it will continue to self-optimize and evolve. For the specific intelligent classification system of Telecommunication network, GSMA, ETSI, and TMForum, have made relevant definitions and suggestions[46,110,111]. The comparison of their classification standards is shown in Table 1.

According to the overview of Telecommunication AI in each Telecommunication ecosystem in Sections 3.1–3.5, and combined with the Level 0–Level 5 (L0–L5) classification system, we have made a grade evaluation on the current application of Telecommunication AI in the field of network infrastructure, network management, communication business, cross-field integration and vertical industry fields, see Table 2.

On the whole, the development and application of Telecommunication AI in various Telecommunication ecosystems is still in the initial stage. In some applications in the field of network management and business support, AI has reached the third level, namely partial autonomy. For example, SON's liberalization and self-healing in the network field and AIOps' automatic business process orchestration and grayscale release in the field of network operation have all been partially autonomous. The intelligent application scenario is clear in Network Management and Operation. It is because they are capable to collect big data and to guarantee computing power enough through server cluster. For example, wireless energy saving, fault detection, etc., inherently have a good foundation and platform for the application of Telecommunication AI, so the development is relatively fast. In addition, AI service capabilities such as intelligent customer service and intelligent marketing involved in the telecommunications business field can well draw on similar application experience in other industries horizontally, so the development is relatively fast, too. However, most of the intelligent processes in other ecological fields is relatively slow at Level 0 or 1. The development of AI in network infrastructure still

Table 1 Internet intelligent grading standards defined in GSMA/ETSI/TMForum.

	GSMA	ETSI	TMForum
Level 0	System provides auxiliary monitoring function Manual execution of dynamic task	Completely manual	Manual O&M
Level 1	Subtask execution according to the existing rules	Network management system generates batch of equipment configuration script	Auxiliary O&M
Level 2	Enabling closed-loop O&M for some units	Achieving partial autonomy	Part of autonomous networks
Level 3	Perceive real-time environment change based on L2 function Optimize and adjust itself in some fields to adapt to the external environment	Conditional autonomy Achieve automated management at some stages of service life cycle	Conditional autonomy network
Level 4	Achieve the prediction of service and customer experience-driven network based on L3 function in more complex cross-domain environment Proactive closed-loop management	High degree of autonomy Achieve the business perception, proactive O&M, and autonomous network driven by self-healing and business based on SLA	Network with high degree of autonomy
Level 5	Own the closed-loop automation function across service, domain and the whole life cycle Complete autonomous network	Complete autonomous management	Complete autonomous network

Table 2 Application grade of telecommunications AI in telecommunication ecosystem.

	Mobile network infrastructure	Network management	Telecommunication service	Cross-domain intelligence convergence	Private network
Level 0	✓	–	–	✓	✓
Level 1	✓	✓	✓	✓	✓
Level 2	–	✓	✓	–	–
Level 3	–	–	–	–	–
Level 4	–	–	–	–	–
Level 5	–	–	–	–	–

needs to be tested by NWDAF, MDAF, RIC, and other actual commercial effects, and also depends on the future openness of 3GPP, O-RAN, ETSI, and other international communication standards organizations to the integration of AI into network infrastructure architecture. The development prospects of AI in cross-domain integration also depend on the effect of tight coupling between AI and core systems or network elements such as CEM and PCF.

4 Next Decade of Telecommunications AI

The next decade of mobile communications will fully evolve to B5G and 6th generation (6G) standards, and it will be a key decade for the deep integration of Telecommunication and AI. Based on the current development of communication technology standards such as 3GPP, ITU-R, ETSI, etc., it is foreseeable that the vision of "complete autonomy" of B5G and 6G will be gradually realized in stages in the future through the deep integration of AI in various fields of the Telecommunication ecosystem.

4.1 Forward-looking view of international telecommunication standard from B5G to 6G

The 6G pre-researches have been started by the European Union, Japan, Repubic of Korea, the United States, and China since 2018. For example, the Hexa-X project was launched by the European Union under the leadership of Nokia[112], and the 6G technology is researched by the Next G Alliance, which is jointly established by American industry companies such as Qualcomm, Microsoft, and Facebook[113]. The Ministry of Industry and Information Technology has expanded the IMT-2020 (5G) promotion group to the IMT-2030 (6G) promotion group. The ITU-T Focus Group on Network 2030 also proposed the Network 2030[114]. At the 34th meeting of ITU Radio Communication Sector Working Party 5D (ITU-R WP5D) on February 19, 2020, in Geneva, Switzerland, a discussion on the 6G research is hold. The discussion includes the research trend and vision of future technology, which forms a timetable of 6G research.

The details on the development of B5G/6G is being planned by various communication standards organizations. According to the white paper[115] and the forecast of the 3GPP and ITU's roadmap to 6G standard, the standardization stage of 3GPP R17–R18 (2020–2023) focuses on researching the trend and vision on 6G technology; the standardization stage of 3GPP R19–R20 (2023–2027) focuses on researching the 6G frequency spectrum and performance; each country will submit its 6G assessment to ITU in the 3GPP R21 standardization stage (2027–2029).

3GPP expects to start the formulation of 6G standard as early as 2025. The formulation of 6G radio interface standard technical specifications will be completed as early as in 3GPP R20 (2026–2027), and the 3GPP 6G standard will be submitted to the ITU in 3GPP R22 (2029–2030). One can expect that B5G/6G will continue evolving and enhancing radio interface protocols in use cases, which includes mobile broadband, fixed wireless access, industrial Internet of Things (IoT), Internet of vehicles, extended reality (XR), large-scale machine-to-machine communication, drones and satellite access. Meanwhile, the related standards for higher frequency bands, such as new radio (NR) in 52.6–71 GHz and terahertz, will be researched and formulated. In addition, the service area by the 6G telecommunication standard will expand from land to satellite, underwater, and underground, which realizes the communication between sea, ground, and sky. For vertical industries such as industrial IoT, the research work on standards will continue. The standards involve cellular narrowband IoT network, mid-range terminals for video surveillance, integrated evolution of access and backhaul, 5G radio interface and its evolution, 5G radio interface in unlicensed spectrum, positioning enhancement, intelligent self-organizing network, communication sensor integration and its evolution, and network topology enhancement. A part of researches mentioned above have started in working groups, for example, 3GPP SA1/SA2[116].

4.2 Forward-looking view of telecommunications AI in mobile network infrastructure

This section will introduce the forward-looking view of AI application in Telecommunication network infrastructure in the next decade from four aspects: wireless access network, core network, transport network, and terminal.

4.2.1 Wireless access network

In the wireless access network, the application scenarios of SON are well defined by 3GPP, and machine learning algorithms applied successfully in some degree in the industry. Therefore, we believe that the development of SON in B5G radio access network will continue accelerating. AI algorithms, such as neural networks and reinforcement learning, will gradually replace traditional machine learning algorithms, for instance genetic algorithms, evolutionary algorithms, and multi-objective optimization algorithms, etc. There will be great potential in the self-optimization and self-healing of SON. Two research topics for 3GPP SA5 and RAN3 are (1) study on SON for 5G, and (2) RAN-centric data collection and utilization for long term evolution and NR[43, 117]. Besides of inheriting most of the scenarios and use cases in the last generation of SON, it is suggested that the next 3GPP's key action for SON is to define the interface and signaling protocols connecting SON to NR, 5GC, OAM, and 4G systems, by which SON can fit 5G network infrastructure as soon as possible. At the same time, 3GPP RAN3 is studying the feasibility of transforming SON to an independent RAN logical entity or function. If this can be achieved, it will be helpful for SON to collect and analyze data on the radio side concurrently, achieving the three core functions of SON, parameter self-configuration, performance self-optimization, and fault self-healing on the radio side.

In addition to the NWDAF defined by 3GPP on the 5G core network side, the EU 5G-MoNArch project team also suggested setting an independent RAN-DAF on the radio side to perform data analysis and decision-making on the CU side of 5G NR[86]. Intelligent analysis on the radio side requires real-time or quasi-real-time decision-making due to its instantaneity, e.g., wireless resource scheduling and management. Therefore, AI analysis on real-time data needs to be performed locally as much as possible to ensure immediate and dynamic optimization for the performance. RAN-DAF, as the AI and data analysis network element of the radio access network, is able to collect and monitor the data of UE and RAN, including channel quality indicator (CQI), power level, path loss, radio link quality, radio resource usage, modulation and coding scheme (MCS), radio link control (RLC) and buffer status information, etc. MoNArch recommends that RAN-DAF sends the above information to the controller on the RAN that is defined by ran controller agent (RCA)-MoNArch. The controller is equivalent to the RIC defined by O-RAN. RAN-DAF and RCA jointly optimize the quality from the perspective of radio side, i.e., elastic slice-aware radio resource control, slice-aware RAT (radio access technology) selection, cross-slice radio resource management, etc. Since RCA is not available in 3GPP, it is practical to set RAN-DAF on the RAN in the form of SBA, and to connect with the core network through a cross-domain message bus. The interface descriptions of RAN-DAF, NWDAF and MDAF are shown in Fig. 15. At present, the RAN-DAF of 5G-MoNArch has not been defined in the 3GPP standard network element, and it is uncertain whether the RAN-DAF will control and manage the RAN, or support some functions of SON in the future. We believe that independent logic virtual functions of either RAN-DAF or SON is enough for applications.

The RIC of O-RAN will continue evolving and strengthening, especially for intelligent policy control for different types of apps. The evolution and enhancement help operators implement service orchestration that is based on App features at the service orchestration layer. The RIC will perceive the type of the app. The RIC uses the characteristics of the app for radio resource management by the third-party xApps through the southbound interface. Based on the type of the app, the northbound interface interacts with the edge server through the network open application program interface (API), as shown in Fig. 16a. RIC's capability exposure and the enhanced radio resource management

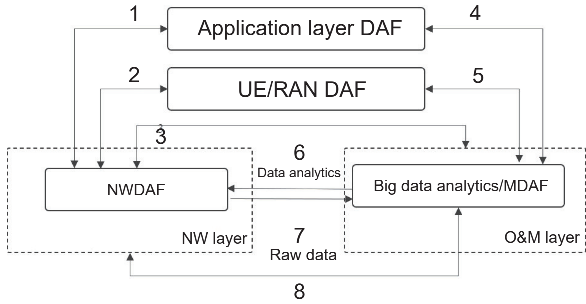


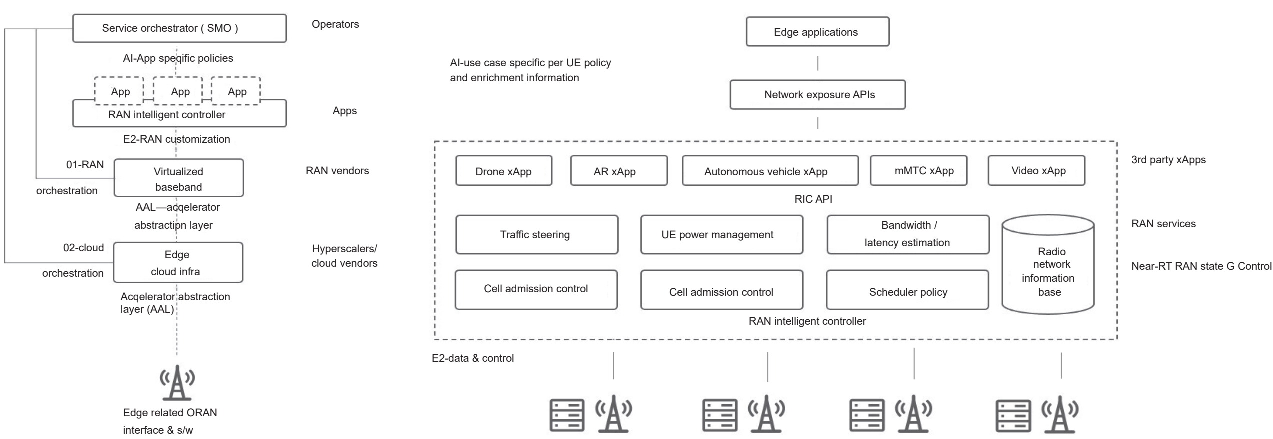
Fig. 15 Data analysis framework defined by 5G-MoNArch.

functions will include, for example, data sharing among multiple O-RAN devices, the guarantee of wireless network slicing for SLA, radio resource optimization for Internet of Vehicles and UAVs (unmanned aerial vehicles), dynamic spectrum sharing, and the combination with MEC to meet the business needs of vertical industries[118], as shown in Fig. 16b. SMO (service management and orchestration) defined by O-RAN can support the non-RT RIC functions and more powerful cloud and wireless networks operations with maintenance management, as the functional architecture and interfaces are mature.

4.2.2 Core network

NWDAF, the AI network element of the core network, will be able to enhance network performance and optimize user experiences in the future, which achieves autonomy and intelligent services within the domain. With the intercommunication with the peripheral network element interfaces of the core network, e.g., other NF (network function), AF (application function), and OAM, etc, and the realization of soft data collection, NWDAF will be able to participate in the decision-making of the core network control plane in a comprehensive and real-time manner[119]. For example, when NWDAF cooperates with network slice selection function (NSSF) and PCF, PCF can make and execute strategy decisions based on the analysis results from the slicing of NWDAF, and NSSF can select the slices based on the load analysis of NWDAF.

A new use case for NWDAF is UE-driven analysis sharing. In this use case, UE information, such as user location information, user profile information, etc., helps NWDAF make intelligent decisions on network slicing. The key function is how to collect and analyze UE information, and how to give the analysis results to other NFs. Other relevant researches include NWDAF-assisted QoS guarantee, traffic processing, personalized mobility



(a) Service orchestration based on O-RAN RIC (b) Capacity openness of O-RAN RIC

Fig. 16 Service orchestration based on O-RAN RIC and capacity openness of O-RAN RIC.

management, policy decision, QoS adjustment, 5G edge computing, NF load balancing, guarantee of SLA slicing, predictable network performance, etc. If the above functions are realized, the intelligence of the 5G core network will be greatly enhanced. Furthermore, the interaction between NWDAF and other fields is worth exploring.

3GPP SA5 is researching on how NWDAF adds the analysis functions to OAM and RAN. In addition, NWDAF will integrate with MEC, and support vertical industry applications through MEC, which empowers more vertical industry applications. In terms of network security, we expect NWDAF will improve functions related to cybersecurity issues. For example, NWDAF will monitor the abnormal behavior of the terminal and network. If any abnormal behavior is observed, NWDAF will report the abnormality to the relevant NF or OAM in time and take protection actions. NWDAF can combine block-chain and other technologies for traceability and security guarantees, e.g., AI models for key information transmission. The converged architecture of NWDAF, MDAF/RAN, and DAF is shown in Fig. 17.

4.2.3 Transport network

With the better integration of SDON/CON and AI, the "zero-touch" cognitive optical transport network (OTN) with fully automated control and management will be realized in the next decade. We can quickly identify transmission problems, predict transmission performance, and optimize transmission parameters, as the OTN-based operation knowledge graph becomes mature. The transmission indicators, such as modulation order, error correction, and wavelength capacity, can be optimally configured using AI technology to ensure transmission performance.

The applications based on IPv6 technology are gradually mature, and AI will play a key role in network routing, SLA assurance of Carrier Network, and deterministic networks to meet customized needs of B5G/6G business scenarios through intelligent IP network from IPv6 to IPv6+.

In terms of cloud-network integration, the computing resources at the edge of the cloud and network will achieve a fully distributed architecture. Seamless and high-quality computing resources are provided based on business needs, which provides guarantees for high-end AI applications at the terminal and edge. A new business model for providing cloud computing services will also appear after the elastic computing power network and dynamic cloud network integration matures. It can use block-chain smart contracts to ensure the security, which solves user privacy issues. The block-chain smart contracts realize the monetization of network and computing resources.

In addition, network slicing is a cross-domain technology that requires efficient cooperation in various fields, such as wireless networks, transport networks, and core networks. Since the transport network is the physical basis for field connection, optimal arrangement and support for transmission resources are very important in the SLA guarantee of slicing. It is expected that AI will gradually mature in the applications of SLA guarantee for end-to-end slicing in transport layer in the future.

4.2.4 Terminal

The connection between the terminal and the network infrastructure interacts with the wireless network through the air interface. AI, which is terminal-based and oriented to future network infrastructure, is mainly applied on the wireless perception of terminals and chips. Equivalently, terminal-based and chip-based AI techniques optimize overhead and latency of wireless access through perception of wireless environment and content. The realization of wireless perception through terminal-based AI is manifested in the following 3 aspects[120]: (1) the spectrum and access perception: a terminal can detect the behavior of other terminals, enhancing 5G system with better access and more efficient orchestration; (2) the content perception: to optimize terminal performance and user experience, the AI infers and analyzes user information, such as location, speed, and mobility, from RF signals, sensors, and traffic behavior data; and (3) the wireless environment perception: new scenarios are generated using the gestures, actions, and certain objects detected by monitoring signal propagation and reflection patterns.

The empowerment of terminal-based AI for future 5G network systems is reflected in three aspects. Firstly, enhancing terminal

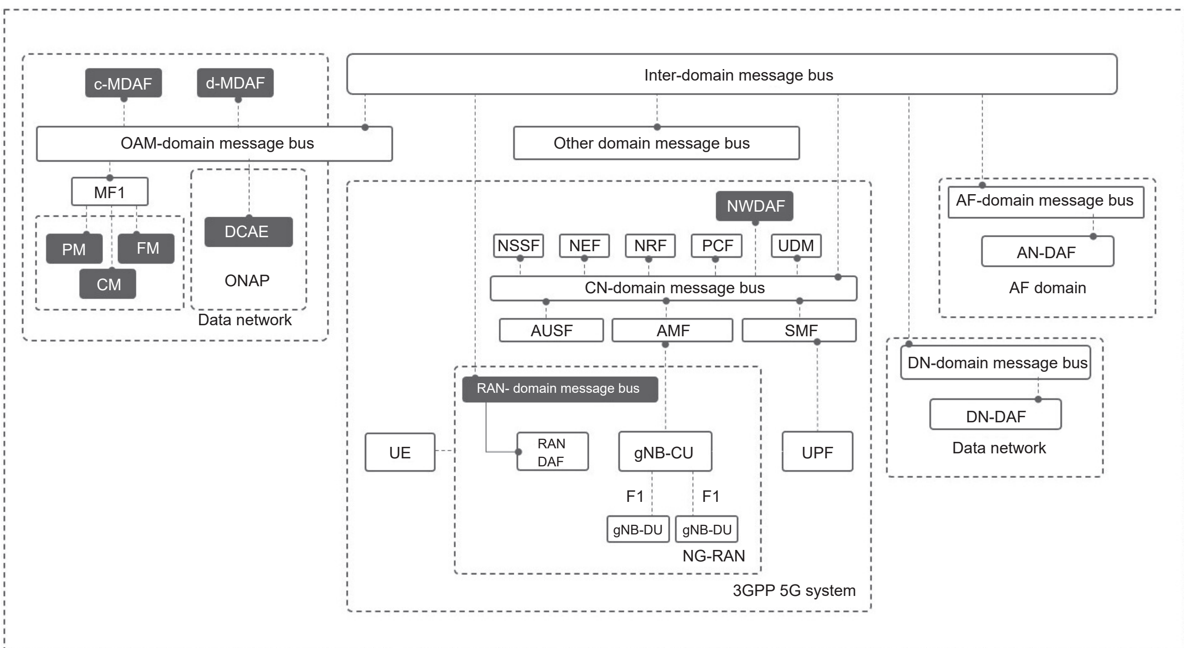


Fig. 17 Forward looking of telecommunications AI in mobile network infrastructure.

experience, intelligent beamforming, and power consumption management can optimize the network speed, the robustness of the network, and the battery life. The intelligent beamforming can be achieved by using deep reinforcement learning, the speed and robustness of network can be improved through the perception of location, velocity, and other parameters in the applications. In terms of energy saving, the AI-based content perception at the terminal can balance the performance and power consumption. Secondly, the improvements of 5G system, which are mainly on the system rate and the spectral efficiency, rely on the location-based wireless interference prediction and Intelligent link adaptation. The load of raw data that needs to be transmitted in the network can be reduced by the assistance of terminal-based AI, which is the intelligent network load optimization. The terminal-centric mobility management can better predict the network switching behavior and timing by terminal AI and sensors, which is intelligent seamless mobility. Thirdly, the AI technology applying at the terminal can instantly detect and defend against malicious base station fraud, malicious interference, and other security hazards, which is wireless security improvement.

4.3 Forward-looking view of telecommunications AI in network management

The development of Telecommunication AI in network management will be carried out from various aspects, such as MDAF, ENI, IBN (intent-based network), and operator's O&M system defined by the ISO. In addition, there will be development in AI network signaling system, network digital twining, and orchestration system.

4.3.1 MDAF

From the perspective of standards, 3GPP will be expected to continue the on-going evolution of MDAF in the SA5 working group, and to enhance the intelligent functions related to operation for management. In terms of SON's intelligence, MDAF has gradually matured in some scenarios, such as coverage enhancement, resource optimization, fault detection, mobility management, energy saving, paging performance management, and SON collaboration. For example, MDAF will provide coverage analysis that is more accurate with the causes of coverage problems. So, it guides the base station to adjust parameters to maintain the user's service experience. MDAF will provide more accurate congestion analysis on RAN user with the causes of the congestion and the relevant policy recommendations. MDAF will provide more accurate resource utilization analysis with strategic recommendations for solving resource utilization problems. For the key parameters of SLA, such as delay and reliability, MDAF will conduct accurate analysis with suggestions on improving experimental performance. In terms of fault management, MDAF will identify fault more accurate with relevant fixing suggestions. MDAF will provide precise policy recommendations on user mobility management to improve the user's handover success rate and network efficiency. Furthermore, it provides more powerful and intelligent services for slicing management, and manages various performance indicators of slices and SLA. In addition, the interaction between MDAF and network devices, such as NWDAF, will be improved.

4.3.2 ETSI ENI

At present, the ENI system defines the functional architecture, but the interface has not been specifically defined yet. We predict that the ENI working group will define related interfaces, which makes ENI more meaningful in terms of deployment and application. In addition, there will be more related works, such as how to combine the data processing mechanism with the ENI architecture, how to deploy ENI in the operator system, how ENI and other intelligent network units work together, how to express and manage intent policies, and flow information telemetry. The definition for relevant scene and functions will be gradually settled. The next version of ENI will enhance in intelligent application scenarios and implementation.

4.3.3 Intent-based networking

At present, 3GPP and ETSI have begun to define IBN[95,121], aiming to define a more intelligent network automation management mechanism. China Telecommunication has carried out the research on IBN. According to the description in 3GPP TR28.812, the intent driven management (IDM) consumer sends an intent request to the IDM provider, and the IDM provider provides the network configuration based on the intent, whose process is shown in Fig. 18. At the same time, IDM provider will monitor the network status and the consumer's satisfaction. If the consumer's intent does not meet, provider will evaluate the intent again with parameter adjustment.

It can be expected that intent-driven management services will continue evolving and maturing in the next decade. The management complexity and the required knowledges for underlying devices are reduced for operators, which increases network management efficiency in multi-vendor applications. Standardization organizations, such as 3GPP SA5 and ETSI ENI, will continue defining reasonable and accurate expression of intent, automation mechanism, and intent lifecycle management. The application scenarios are expected to be gradually enhanced in terms of network service provisioning, optimization for slicing resource utilization, slicing quality assurance, network optimization, network capacity management, and network function deployment.

4.3.4 Network AI signaling system

The AI platform of the 5G OSS network middle-office system can be regarded as an AI engine for network management and operation. The platform needs to communicate with data collection network elements or module in southbound interface, as well as business systems, i.e., network orchestration, network performance, network resources, network failures, in the northbound interface of 5G OSS through a standard command system. We define this standard command system as network AI signaling system.

Network AI Signaling System is different from the 4G and 5G network signaling systems, which focus on the interconnection and interaction management among network elements. Network AI Signaling System is a standard interconnection and AI

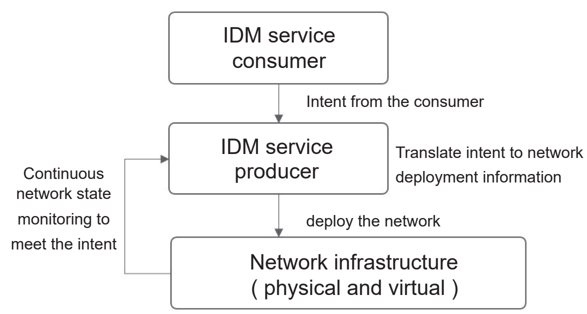


Fig. 18 Intention management.

interactive management command system that connects AI network middle-office or platform, north-south interfaces, network elements, OAM modules, and OSS systems. It includes the definition of commands for AI network management, the upload and distribute mechanism of AI command for flows between interfaces, the execution mechanism of AI analysis for network, AI network authentication, AI network authorization. Also, it includes AI algorithms' training, inference, publishing, deployment, calculation, as well as, the storage resource management, monitor and security of AI network commands, etc.

The typical AI network signaling interfaces are shown in Fig. 19, including the data collection and command execution interfaces of the AI network middle-office and the 3GPP SON system, 3GPP NWDAF, O-RAN RIC, ETSI ENI systems, and 5G OSS systems oriented by the AI network middle-office, which include network orchestration, network performance, network resources, network failures, etc. Figure 20 shows a typical AI network signaling packet structure and the schematic diagram of AI network signal process for abnormal detection of cell telephone traffic.

4.3.5 Network orchestration

Under the trend of software-defined and cloud-based networks, NF management will be taken over by software-defined hypervisors, and it will transform from proprietary hardware-oriented to virtualized management of shared computing and communication resource pooling. Therefore, both function management and network orchestration of the traditional relatively static OSS/BSS will play a crucial role in the future evolution of network technologies. In 2014, ETSI published the management and orchestration standard specification of network functions virtualization (NFV), which initially defined the functional architecture for MANO in an NFV environment[122]. With the built-in of AI for network orchestration, it is necessary for the industry to determine the evolution of network orchestration in 5G era. Meanwhile, the industry needs to clarify the logical and physical connection between AI and network orchestration. As shown in Section 3.2, network orchestration is very important in an operator's network management system. The main functions—network connection and construction, network resource scheduling and orchestration, network workflow and business requirement translation—are network (topology) orchestration, network resource orchestration, and network service orchestration respectively. In 5G OSS, network service orchestration could be an independent subsystem, which is mainly responsible for the orchestration and management of network slicing services that are composed of 5G virtual network functions (VNFs). SLA-based slicing intelligent management will be the main application scenario of communication using AI in network service orchestration. Network services, such as the guarantee and optimization of slice quality, involve the immediate and intelligent orchestration support from the underlying network to upper layer services. At this time, the network resource orchestration plays the role in guaranteeing and supporting the upper layer network service orchestration. The network resource orchestration needs to achieve three elasticities in the future: (1) computing elasticity, i.e., computing is elastic when designing and scaling VNFs; (2) the elasticity of orchestration by flexibly setting VNFs; and (3) the cross-slice resource supply mechanism achieves slice-aware elasticity[123]. The basic network resource orchestration has been defined in ETSI MANO. In terms of how to realize the joint orchestration between network resources and network services based on SLA or ELA (experience level agreement) in the future, and the clarification for the logic and physical connection between AI and network orchestration, ETSI ENI currently provides a reference architecture for the industry to continue developing[123], as shown in Fig. 21.

At the same time, before the existing traditional network with poor flexibility and weak automation is fully replaced, how to coexist and work together with the new network that supports the intelligent orchestration of network topology, resources, and services is a problem for the industry, which needs to form the goal for coherent action as soon as possible. For example, for the traditional network services, such as private lines, how to make rules for the orchestration processes so that they are automated

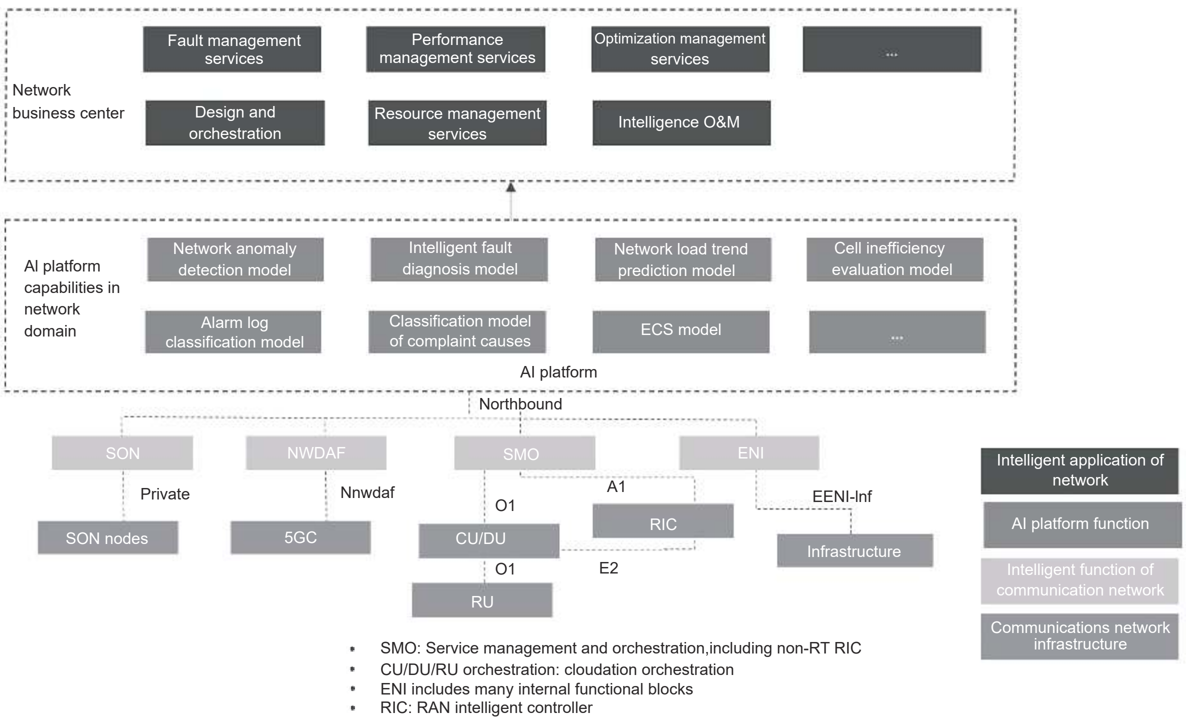
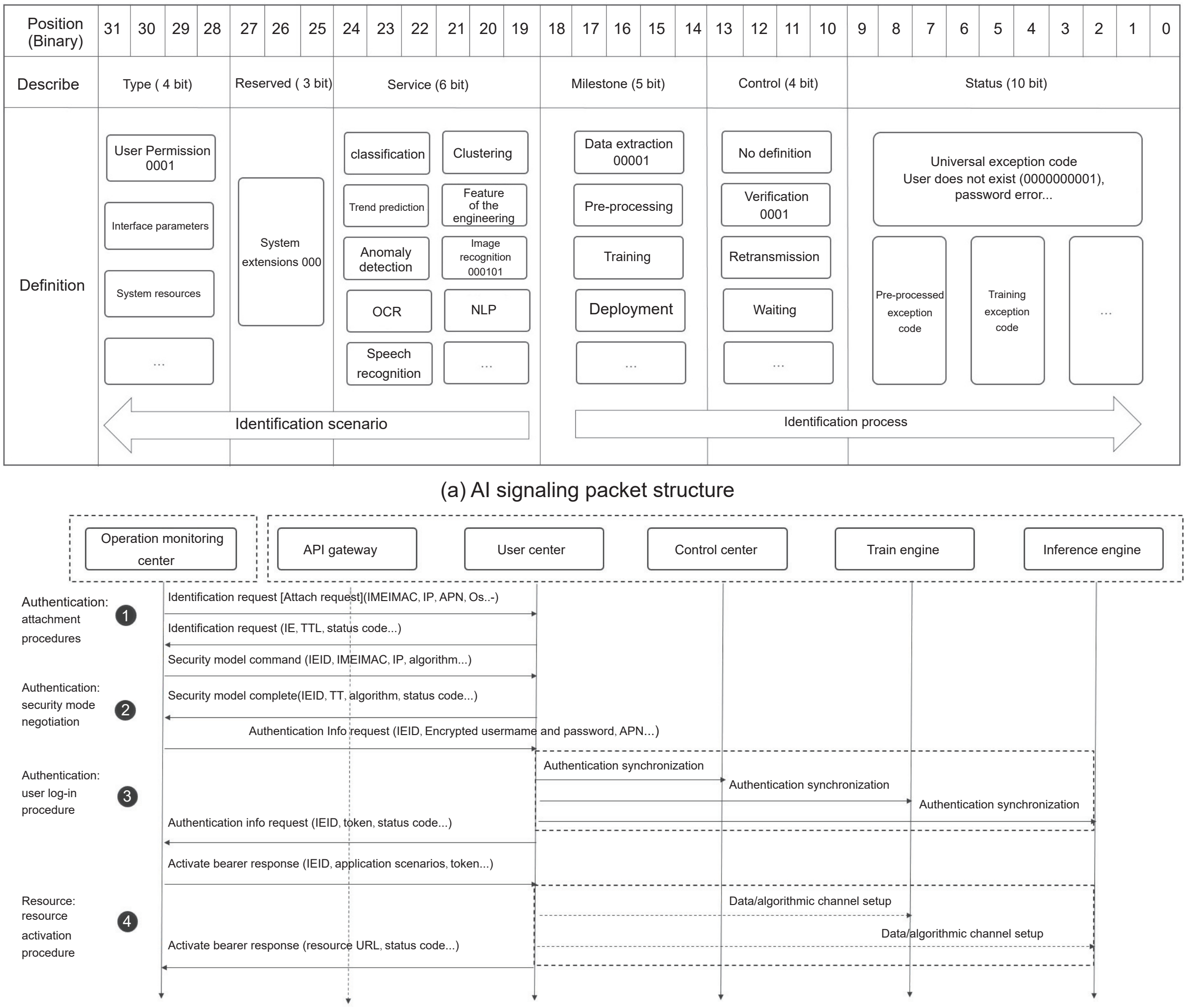


Fig. 19 Network AI signaling interface.



(a) AI signaling packet structure

(b) Procedure of network AI signaling process for abnormal detection of cell traffic load

Fig. 20 Typical network AI signaling packet structure and the application for abnormal detection.

and relatively standardized and they can be integrated with the new 5G network orchestration system?

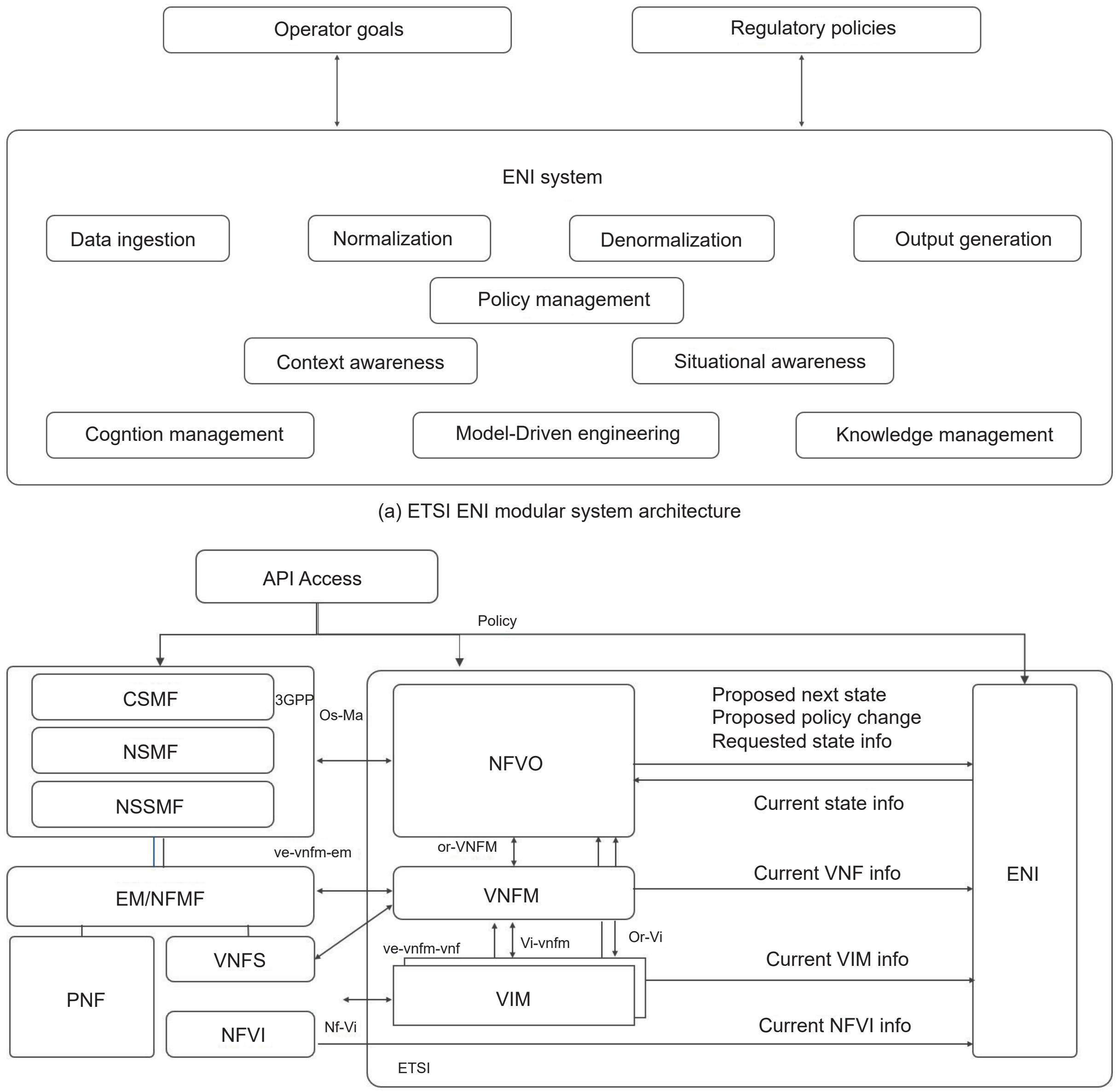
At present, the network automation and intelligent orchestration of global operators are still in the initial stage, and they need to be further improved at both technical and standard levels. It can be expected that, with the deep integration of AI into network orchestration systems, the network (topology) orchestration, network resource orchestration, and network service orchestration will be continuously improved. In particular, to optimize SLA or ELA, AI will play an important role in the translation of network business requirements, and the active and elastic matching between network topology construction and resource assurance. At the standard level, it is expected that standardization organizations, such as 3GPP and ETSI, will gradually improve the definition of relevant scenarios, interfaces, and processes.

4.4 Forward-looking view of telecommunications AI in the field of telecommunication business

For the field of BSS intelligence/management finance, it can be expected that AI will be fully empowered in customer management, package recommendation, and financial intelligence management in the next 10 years, and the transition from primary level to advanced level will be realized. It will play a key role from the establishment of a people-oriented comprehensive customer experience, to the establishment of a more efficient business operation process for telecommunication operators, finally to the rapid innovation and realization of new services, new models, and new technologies. The following will be explained from three perspectives: customer operation, business operation, business model, and operation model.

4.4.1 Customer operation

From the perspective of customer operation, telecommunication operators have completed the comprehensive customer experience stage upgradation from customer-centric to experience-centric. Telecommunication operators focus on operations refinement and service satisfaction improvement in the process of providing customers with marketing, sales, and service. In addition, they take the customer experience in the process of network and business use as the important indicators and join the comprehensive customer experience system for unified evaluation. A comprehensive customer-centric journey experience is established by collecting, aggregating, correlating, and mining customer experience in the process of channel contact, network service, and business use. On this basis, with the assistance of AI algorithms, the needs of customers in more segmented scenarios are met, and the intelligent and automated interaction capabilities are further improved to form a customer-centric comprehensive customer experience at the same time.



(a) ETSI ENI modular system architecture

(b) Management and orchestration architecture

Fig. 21 ETSI ENI architecture and its interaction with ETSI NFV MANO framework.

4.4.2 Business operation

From the perspective of business operations for telecommunication operators, they have basically completed the end-to-end digital upgradation of the process. They are completing the intelligence of business processing and further improving business operation efficiency by implanting big data and AI into existing business operation processes. With the introduction of technologies such as RPA and intelligent business process management suite (IBPMS), it is expected that the business process of manual intervention will be further reduced in the future, the efficiency of process operation will be improved, and the cost will be reduced. With the introduction of AI into the operator's risk control system, the revenue guarantee capability will be improved, and the risk of arrears will be reduced. Operators can carry out more innovative businesses based on their own risk control expectations, which will create the healthy and upward cash flow for telecommunication operators.

4.4.3 Business mode and operation mode

For the current business model and operation model of telecommunication operators, the method of investing in a heavy-asset model and using volume-based pricing in the mass market has been facing difficulty to meet the needs of rapid development. Furthermore, the traditional manual method of product formulation and pricing model formulation that using large-scale segmented customer groups for the mass market is also difficult to meet the requirements of future user scenarios and personalization. Therefore, operators need to make full use of AI capabilities, use virtual experts/personal assistants to subdivide and operate more fine-grained markets, cooperate with intelligent workflow and intelligent risk control system, and promote "one customer, one policy" method to further enhance innovation through creative business products and operating models.

4.5 Forward-looking view of telecommunications AI in cross-domain integration of intelligence

For the cross-domain intelligence in networks and business, it is expected that the architecture and functions of CEM and PCF+ will continue to develop in the next 10 years, and the customer experience perception system will evolve from SLA to ELA.

4.5.1 Enhancement of customer experience perception and evolution from SLA to ELA

Since the 1990s, the QoS system[124,125] defined by international standard organizations such as ITU and ISO has been adopted by the vast majority of telecommunication operators and used in the SLA signed with their users. Traditional QoS is driven by technology, especially the QoS defined by network and service performance indicators[126,127], that is, the network indicators are incorporated into SLAs. Since traditional QoS cannot truly reflect users' experience and feelings when using network services, the QoS system has gradually developed and evolved to a user-centric QoE system in recent years. ITU defines QoE as the overall acceptability of an application or service that can be subjectively perceived by end users[124].

Under the current SLA system, the service level provided by telecommunication operators to users is "best effort". For example, a bandwidth of 100 Mbit/s usually means "up to 100 Mbit/s ideally". This kind of best-effort service and KPIs from a technical perspective are difficult to directly relate to the user's QoE. At the same time, when the user pays for the service, they always have an expectation that the stability and quality of the service should be equal to or greater than the cost of payment. As CEM's QoE algorithm system gradually matures and improves in the future, telecommunication operators will be able to use the QoE algorithm system to predictably evaluate and proactively manage user experience expectations. Therefore, it is also imperative for the traditional user-oriented SLA system to evolve to the QoE-based ELA system. ELA is defined as "a special SLA, that is, a consensus based on quality levels formed by user experience of using a service on the premise that the customer understands a certain service clearly"[120].

Figure 22 depicts the corresponding relationship of QoS guarantee among telecommunication operators, content providers,to business (ToB) users, and to customer (ToC) users[120]. ELA is mainly for ToC users, operators, and content providers. In the current business model, ELA is primarily a hybrid model of web services plus a content quality-based model, that is a whole set of ELA system includes the QoS of network and the QoE of application layer. Currently, it is difficult for users to clearly define whether the decline in experience is on the network side or the application side. The future ELA system should be constructed based on Fig. 22b, that is, build an separate ELA system on facing the network side and the application side for ToC users. For users, in the area of QoE perception, breaking down the quality-of-experience black box that is attributed to "inadequate network" can make it clearer that the quality of experience provided by different service providers has its own advantages and disadvantages. At the same time, in order to provide better QoE guarantee for ToC customers, content providers can also define various QoS requirements more clearly in the SLA with the communication operator Internet service provider (ISP).

4.5.2 Evolution of CEM architecture

Because CEM data comes from multiple domains: network OSS domain, business BSS domain, and application layer. "Data compliance" and "data silos" make it extremely challenging for CEM to aggregate data from multiple domains[128]. In order to balance the dual requirements of "privacy security" and "data fusion", the idea of constructing a CEM model based on cross-regions is proposed. There are two most common methods for achieving cross-regional modeling: transfer learning (TL) and federated learning (FL). Through large-scale training of massive data, a pre-trained model (PM) with strong universality is obtained, and then the pre-trained model is moved to the application scenario with small-scale data, and fine-tune pre-trained models with small-scale data to achieve significant improvements in model performance[129]. This idea can be used to build a CEM model. As shown in Fig. 23a, the basic PM of CEM is obtained through data training of a communication operator enterprise, and then the PM is fine-tuned by using the scene business data to obtain a scene-based CEM, and avoid the direct fusion of data.

In 2017, a federated modeling method for fragmented data, i.e., federated learning, was proposed[25]. Federated learning is very suitable for CEM modeling scenarios. As shown in Fig. 23b, federated learning adopts the idea of "data do not move, model moves", which avoids the privacy and security issues about users' experience data. Throughout the joint modeling process, the data of all parties are not released from the database and are kept locally. Therefore, the user experience perception method based on federated learning technology is safe, controllable, and stable, which greatly improves the accuracy of the joint model and the comprehensiveness of the assessment.

4.5.3 Evolution of PCF+

For the subsequent evolution version of PCF+, it can be expected that PCF+ will provide users with more accurate, real-time, and differentiated policy control through the interaction with the OSS domain and the BSS domain, as shown in Fig. 13. PCF+ can unify the BSS domain policy and ensure the reliability and real-time performance of the service during the process of providing policy control. For the policy control on the network side, PCF+ can

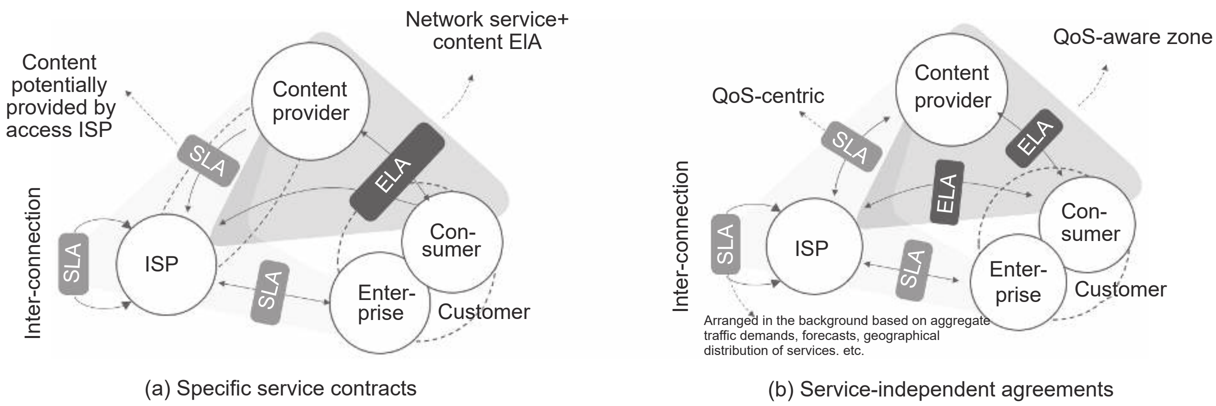
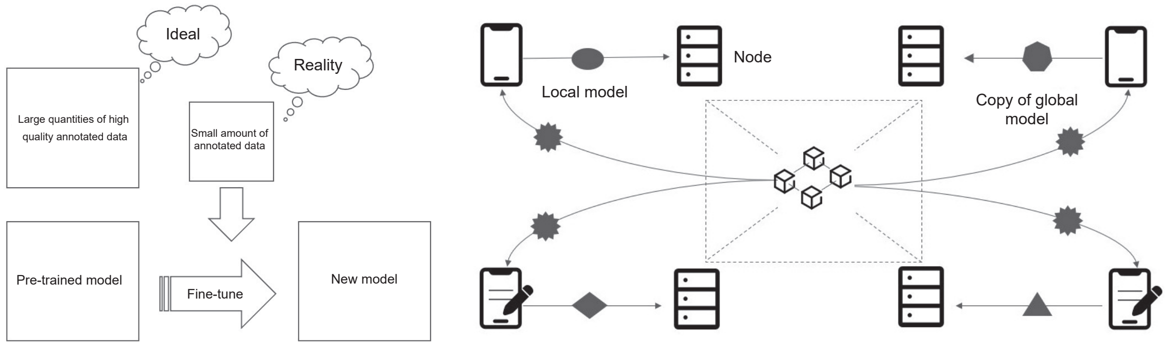


Fig. 22 Construction of ELA ecology through SLA and QoS concept.



(a) CEM modeling based on transfer learning[129] (b) CEM modeling based on federated learning[25]

Fig. 23 CEM modeling methods based on transfer learning and federated learning.

dynamically control user QoS parameters and session management in real time to ensure service quality. Because PCF+ can pull through the OSS/BSS domain, the policy control of users becomes more personalized, which can further explore scenarios with added value and realize the business vision of B5G/6G.

From the perspective of the deployment method of PCF+, it can be deployed in the OSS domain or in the BSS domain to provide services targeted at different domains. PCF+ between different domains can interact with each other to obtain their data information. PCF+ can also be deployed in the hierarchical architecture according to the actual network conditions.

4.6 Forward-looking view of telecommunications AI in private network applications

One of the core values of 5G is the proprietary applications for enterprise ToB. It is expected that in the next 10 years, in vertical fields (such as the Internet of Vehicles, Intelligent Manufacturing, High-Definition Video/VR/AR (augmented reality), Telemedicine, and Smart Cities), telecommunication AI will help enterprises to achieve advanced intelligence and even fully intelligent private network functions.

With the way things are, although there have been some deployment cases of 5G private networks in the industrial Internet field, and such cases have effectively promoted the application and ecological construction of 5G in vertical industries and meet diverse network requirements with different deployment forms of private networks, there is still a big gap to reach the requirements of vertical industries. It can be expected that in the next 10 years, telecommunication AI can fully meet the requirements of vertical industries for high-quality communication and network security through the combination of MEC and business as well as the maturity of algorithms. For example, cutting-edge AI technologies such as federated learning and transfer learning will help private networks solve problems such as data security, privacy, and lack of data volume. AI technology will be used to perceive business changes, optimize wireless network parameters to ensure business transmission quality, and truly transform private networks into a high-performance, secure, and reliable private network.

5 Expectable Future: Comprehensive Promotion of Telecommunication Intelligentization in the Next Decade

For the process of 6G evolution, 3GPP follows the speed of one Release every 2 to 3 years, and it is expected to evolve from 3GPP R17 to 3GPP R21 in the next 10 years. At present, ITU-R has started research on 6G technology trends and visions, and is expected to formulate a formal standard specification for 6G in 2027–2028. As shown in Fig. 24, during the evolution of 3GPP and ITU-R technical standards, various functional areas of mobile communication (core network, wireless, transmission, network management, business support, network application, etc.) will gradually reach the high-level stage of intelligence in the communication ecosystem in the next 10 years through different

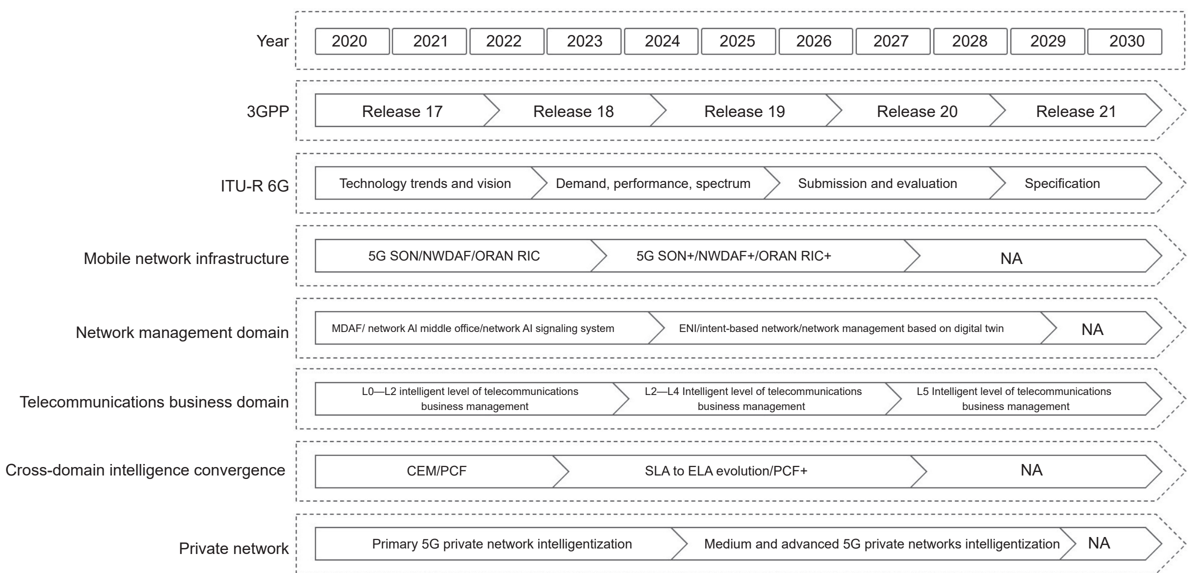


Fig. 24 Forward looking of telecommunications AI in the next decade.

degrees of coupling with AI, which will eventually achieve the goals of communication autonomy and full intelligence in the B5G/6G vision.

AI is oriented to the telecommunication network infrastructure. In the next 3–5 years, NWDAF will gradually become mature and commercialized in the 5G core network. Wireless and core network optimization will also use SON to achieve the goal of AI-driven intelligent network optimization. Commercial deployment of SON will be possible as an independent SON system or integrated into the 5G OSS system. Whether RAN-DAF will be defined as an independent network element has not yet been determined. In the next 5 to 10 years, with the gradual commercialization of O-RAN, RIC will also be commercially deployed as an intelligent controller for open wireless networks.

AI is oriented to network management and operation. In the next 3–4 years, MDAF will implement part of data analysis functions at the network management level. With the construction of the network middle-office, the network AI middle-office will realize commercial deployment in the 5G OSS of some operators. The network AI signaling system, as an interactive language of AI and network, will empower AI into the network ecosystem. In the next 5—10 years, IBNs and the ETSI ENI network experience and perception system will gradually mature, and will be applied in the network infrastructure architecture in the mid-to-late 5G period. The digital twin technology will be combined with network simulation and AI to realize twinning and intelligent management on the planning, construction, optimization, and maintenance of the entire network life cycle.

AI is oriented to telecommunication business and support, and the technology middle-office system built by some telecommunication operators will be fully commercialized and mature in the next 3–5 years. The AI platform sector will serve as a carrier for AI to inject intelligence and empowerment in the BSS field, and comprehensively promote the intelligence into customer operations and business operations. Due to the similar application experience in vertical industries, some sub-fields of telecommunication business (such as intelligent customer service, intelligent marketing, intelligent recommendation, etc.) will accelerate their development in the next 5–10 years, and may reach the highly intelligent level of L4 or L5 earlier.

AI is oriented to the integration of intelligence across BSS and OSS domains, and CEM and PCF will develop along the evolution path of the integration of BSS and OSS domain. CEM will combine network and business data to achieve closed-loop management of customers' network and business experience perception in the customer's full life cycle journey. Due to the use of cross-domain data, the future architecture of CEM can be achieved by opting for federated learning. The customer experience and perception management system will evolve from SLA to ELA system. PCF can provide accurate, real-time, and personalized policies and services for the network, business, and customers through the interaction between the OSS domain and the BSS domain.

The vertical industry private network will be in the initial stage of commercial construction in the next 3–4 years, and the main deployment mode will be virtual private network. Therefore, the application of AI for virtual private networks will focus on the SLA guarantee of 5G private network slices, the intelligent scheduling, optimization of slice resources, wireless private network coverage, and performance optimization. In the next 5–10 years, hybrid private networks and independent private networks will gradually be deployed and developed. The application of AI to independent or hybrid private networks will focus on accurate QoS guarantee for ToB services, service experience perception and real-time evaluation for ToC services, and AIOps. In addition, through the interaction with the private network application platform multi-access edge platform (MEP) of vertical industries, AI technology can adaptively adjust the parameter settings on the application layer to guarantee the QoS of edge applications. In the primary intelligent stage of the industry private network, AI is mainly used for the intelligence of performance, quality, operation and maintenance guarantee. In the intermediate and advanced intelligent stage, it is used for intelligent control and management of high reliability, low latency, and multiple concurrent connections.

In the future, the application of telecommunication AI systems will be further enhanced in terms of security, robustness, and interpretability. In particular, the technical combination of federated learning, block-chain, and privacy computing in AI is expected to be developed in various telecommunication ecosystems to solve the problem of data silos and privacy security between the telecommunication ecosystem and vertical industries. Since the federated learning model involves the cooperative training with multi-party data, the federated center is responsible for key management and model gradient management, so the federated center needs to be audited regularly. Block-chain technology solves the problem of consensus and credibility, so model training, inferencing, and character alignment are all managed through block-chain network. Thus the trusted network of multi-party cooperation is realized through smart contracts and consensus computing, and the role of the central node can be replaced by the block in the case of multi-party federation.

For 6G, telecommunication AI will promote the intelligent integration of the multi-dimensional ecosystem of air, sky, land and sea. Due to the expansion of the 6G telecommunication ecosystem in the spatial dimension, more scenarios will emerge. Telecommunication AI will solve non-deterministic polynomial-time hard (NP-hard) problems such as cross-layer optimization and joint optimization of multiple ecological subsystems.

At present, the architectural integration and functional application of AI in the telecommunication ecosystem have been standardized and defined by 3GPP, ITU-R, and ETSI. The commercial process of communication AI in 5G is in the early stage, especially the network elements related to communication AI (such as 3GPP NWDAF or O-RAN RIC) are in the testing stage, so they are rarely commercialized in 5G networks. Currently, there are many applications of telecommunication AI for network management, business management and application layer, and staged results have also been achieved. We call on the communication industry builders to further open the network standardization interface, give AI opportunity to fully empower and focus on the network infrastructure, network management and operation system, and business support system, so that the potential of combining 5G and AI as a general-purpose technology is fully unleashed within the communications ecosystem and in vertical applications.

Dates

Received: 29 November 2021; Revised: 25 April 2022; Accepted: 22 June 2022

References

[1] R. Li, Z. Zhao, X. Zhou, G. Ding, Y. Chen, Z. Wang, and H. Zhang, Intelligent 5G: When cellular networks meet artificial intelligence, *IEEE Wirel. Commun.*, vol. 24, no. 5, pp. 175–183,

2017.

[2] G. E. Hinton, S. Osindero, and Y. W. Teh, A fast learning algorithm for deep belief nets, *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, 2006.

[3] D. S. Landes, *The Unbound Prometheus*: *Technological Change and Industrial Development in Western Europe from 1750 to the Present*. Cambridge, UK: Cambridge University Press, 1969.

[4] N. Rosenberg, *Inside the Black Box*: *Technology and Economics*. Cambridge, UK: Cambridge University Press, 1983.

[5] T. F. Bresnahan and M. Trajtenberg, General purpose technologies "Engines of growth"? *J. Econom.*, vol. 65, no. 1, pp. 83–108, 1995.

[6] R. G. Lipsey, K. I. Carlaw, and C. T. Bekar, *Economic Transformations*: *General Purpose Technologies and Long-Term Economic Growth*. Oxford, UK: Oxford University Press, 2006.

[7] The Bureau of Communication and Arts Research, Impacts of 5G on Productivity and Economic Growth, https://www.infrastructure.gov.au/sites/default/files/impacts-5g-productivity-economic-growth.pdf, 2018.

[8] Ministry of Internal Affairs and Communications, https://www.soumu.go.jp/english/index.html, 2019.

[9] J. E. Prieger, An Economic Analysis of 5G Wireless Deployment: Impact on the U. S. and Local Economies, https://actonline.org/wp-content/uploads/ACT-Report-An-Economic-Analysis-of-5G-FINAL.pdf, 2020.

[10] D. J. Teece, 5G mobile: Disrupting the automotive sector, https://www.qualcomm.com/content/dam/qcomm-martech/dm-assets/documents/5g_disrupting_automotive_drteece_paper_may2017_final.pdf, 2017.

[11] J. E. Smee and J. Hou, 5G+AI: The ingredients fueling tomorrow's technology innovations, https://www.qualcomm.com/news/onq/2020/02/5gai-ingredients-fueling-tomorrows-tech-innovations, 2020.

[12] R. Shafin, L. Liu, V. Chandrasekhar, H. Chen, J. Reed, and J. C. Zhang, Artificial intelligence-enabled cellular networks: A critical path to beyond-5G and 6G, *IEEE Wirel. Commun.*, vol. 27, no. 2, pp. 212–217, 2020.

[13] J. McCarthy, M. L. Minsky, N. Rochester, and C. E. Shannon, A proposal for the Dartmouth summer research project on artificial intelligence, August 31, 1955, *AI Mag.*, vol. 27, no. 4, pp. 12–14, 2006.

[14] A. L. Samuel, Some studies in machine learning using the game of checkers, *IBM J. Res. Dev*., vol. 44, no. 1.2, pp. 206–226, 2000.

[15] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, Learning representations by back-propagating errors, *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[16] L. N. Yasnitsky, Whether be new "winter" of artificial intelligence? in *Integrated Science in Digital Age*, T. Antipova, Ed. Cham, Germany: Springer, 2019, pp. 13–17.

[17] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ImageNet classification with deep convolutional neural networks, in *Proc. 25th Int. Conf. on Neural Information Processing Systems*, Lake Tahoe, NV, USA, 2012, pp. 1097–1105.

[18] A. Graves, A. Mohamed, and G. Hinton, Speech recognition with deep recurrent neural networks, in *Proc. 2013 IEEE Int. Conf. Acoustics, Speech and Signal Processing*, Vancouver, Canada, 2013, pp. 6645–6649.

[19] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, Distributed representations of words and phrases and their compositionality, in *Proc. 27th Annu. Conf. Neural Information Processing Systems*, Lake Tahoe, NV, USA, 2013, 3111–3119.

[20] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, Generative adversarial nets, in *Proc. 27th Int. Conf. on Neural Information Processing Systems*, Montreal, Canada, 2014, pp. 2672–2680.

[21] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, et al., Human-level control through deep reinforcement learning, *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[22] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, in *Proc. 2019 Conf. North American Chapter of the Association for Computational Linguistics*: *Human Language Technologies*, Minneapolis, MN, USA, 2019, pp. 4171–4186.

[23] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, et al., Language models are few-shot learners, in *Proc. 34th Conf. on Neural Information Processing Systems*, Vancouver, Canada, 2020, pp. 1877–1901.

[24] J. Konečnỳ, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, and D. Bacon, Federated learning: Strategies for improving communication efficiency, arXiv preprint arXiv: 1610.05492, 2016.

[25] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. Arcas, Communication-efficient learning of deep networks from decentralized data, in *Proc. 20th Int. Conf. on Artificial Intelligence and Statistics*, Fort Lauderdale, FL, USA, 2017, pp, 1273–1282.

[26] Q. Yang, Y. Liu, T. Chen, and Y. Tong, Federated machine learning: Concept and applications, *ACM Trans. Intell. Syst. Tech*., vol. 10, no. 2, p. 12, 2019.

[27] Y. Okumura, Field strength and its variability in VHF and UHF land-mobile radio service, *Rev. Electr. Commun. Lab.*, vol. 16, pp. 825–873, 1968.

[28] M. Hata, Empirical formula for propagation loss in land mobile radio services, *IEEE Trans. Veh. Technol.*, vol. 29, no. 3, pp. 317–325, 1980.

[29] *Radio Frequency* (*RF*) *system scenarios*, 3GPP TR 25.942, 1999.

[30] D. J. Young and N. C. Beaulieu, The generation of correlated Rayleigh random variates by inverse discrete Fourier transform, *IEEE Trans. Commun.*, vol. 48, no. 7, pp. 1114–1127, 2000.

[31] C. X. Wang and M. Patzold, Methods of generating multiple uncorrelated Rayleigh fading processes, in *Proc. 57th IEEE Semiannual Vehicular Technology Conf*., Jeju, Republic of Korea, 2003, pp. 510–514.

[32] A. Aragón-Zavala, B. Belloul, V. Nikolopoulos, and S. R. Saunders, Accuracy evaluation analysis for indoor measurement-based radio-wave-propagation predictions, *IEE Proc. -Microwaves, Antennas Propag.*, vol. 153, no. 1, pp. 67–74, 2006.

[33] *Telecommunication management*; *Self-organizing networks* (*SON*); *Concepts and requirements*, 3GPP TS 32.500, 2008.

[34] *Telecommunication management*; *Self-configuration of network elements*; *Concepts and requirements*, 3GPP TS 32.501, 2008.

[35] J. M. Johnson and V. Rahmat-Samii, Genetic algorithms in engineering electromagnetics, *IEEE Antennas Propag. Mag.*, vol. 39, no. 4, pp. 7–21, 1997.

[36] O. Boyabatli and I. Sabuncuoglu, Parameter selection in genetic algorithms, *J. Syst., Cybern. Inf.*, vol. 4, no. 2, p. 78, 2004.

[37] J. Pérez-Romero, O. Sallent, R. Ferrús, and R. Agustí, Knowledge-based 5G radio access network planning and optimization, in *Proc. 2016 Int. Symp. on Wireless Communication Systems*, Poznan, Poland, 2016, pp. 359–365.

[38] *Study of enablers for network automation for 5G*, 3GPP TR 23.791, 2018.

[39] ETSI, Improved Operator Experience Through Experiential Networked Intelligence (ENI), https://www.etsi.org/images/files/ETSIWhitePapers/etsi_wp22_ENI_FINAL.pdf, 2017.

[40] China Communications Standards Association, https://ccsa.org.cn/standardDetail?standardNum=SR%20290-2020, 2020.

[41] O-RAN ALLIANCE, O-RAN Use Cases and Deployment Scenarios, https://assets-global.website-files.com/60b4ffd4ca081979751b5ed2/60e5aff9fc5c8d496515d7fe_O-RAN%2BUse%2BCases%2Band%2BDeployment%2BScenarios%2BWhitepaper%2BFebruary%2B2020.pdf, 2020.

[42] *Study on RAN-centric data collection and utilization for LTE and NR*, 3GPP TR 37.816, 2018.

[43] *Telecommunication management*; *Study on the self-organizing networks* (*SON*) *for 5G networks*, 3GPP TR 28.861, 2018.

[44] *Telecommunication management*; *Study on management aspects of communication services*, 3GPP TR 28.805, 2018.

[45] *Architectural framework for machine learning in future networks including IMT-2020*, ITU-T Y.3172, 2019.

[46] GSMA, AI in Network Use Cases in China, https://www.gsma.com/greater-china/wp-content/uploads/2019/10/AI-in-Network-Use-Cases-in-China.pdf, 2019.

[47] *Study on concept*, *requirements and solutions for levels of*

autonomous network, 3GPP TR 28.810, 2019.

[48] SA WG2 Meeting #139E, https://www.3gpp.org/ftp/tsg_sa/WG2_Arch/TSGS2_139e_Electronic/Inbox/Chairmans_Notes/Chairmans Notes_Tao_Ph1_06-04-1600.doc, 2020.

[49] D. Poole, A. Mackworth, and R. Goebel, *Computational intelligence: A logical approach*. Oxford, UK: Oxford University Press, 1998.

[50] S. Russell and P. Norvig, *Artificial Intelligence*: *A Modern Approach*. 3rd ed. Hoboken, NJ, USA: Pearson Education, 1995.

[51] S. S. Mosleh, L. Liu, C. Sahin, Y. Zheng, and Y. Yi, Brain-inspired wireless communications: Where reservoir computing meets MIMO-OFDM, *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 10, pp. 4694–4708, 2018.

[52] N. Samuel, T. Diskin, and A. Wiesel, Deep MIMO detection, in *Proc. 18th Int. Workshop on Signal Processing Advances in Wireless Communications*, Sapporo, Japan, 2017, pp. 1–5.

[53] F. Carpi, C. Häger, M. Martalò, R. Raheli, and H. D. Pfister, Reinforcement learning for channel coding: Learned bit-flipping decoding, in *Proc. 57th Ann. Allerton Conf. on Communication*, *Control*, *and Computing*, Monticello, IL, USA, 2019, pp. 922–929.

[54] W. Lyu, Z. Zhang, C. Jiao, K. Qin, and H. Zhang, Performance evaluation of channel decoding with deep neural networks, in *Proc. 2018 IEEE Int. Conf. on Communications*, Kansas City, MO, USA, 2018, pp. 1–6.

[55] K. Yashashwi, D. Anand, S. R. B. Pillai, P. Chaporkar, and K. Ganesh, MIST: A novel training strategy for low-latency scalable neural net decoders, arXiv preprint arXiv: 1905.08990, 2019.

[56] F. Liang, C. Shen, and F. Wu, An iterative BP-CNN architecture for channel decoding, *IEEE J. Sel. Top. Signal Process.*, vol. 12, no. 1, pp. 144–159, 2018.

[57] *Evolved universal terrestrial radio access network* (*EUTRAN*); *Self-configuring and self-optimizing network* (*SON*) *use cases and solutions*, 3GPP TR 36.902, 2008.

[58] *Evolved universal terrestrial radio access* (*E-UTRA*); *Potential solutions for energy saving for E-UTRAN*, 3GPP TR 36.927, 2010.

[59] Z. Lin, Y. Ouyang, L. Su, W. Lu, and Z. Li, A machine learning assisted method of coverage and capacity optimization (CCO) in 4G LTE self organizing networks (SON), in *Proc. 2019 Wireless Telecommunications Symp.*, New York, NY, USA, 2019, pp. 1–9.

[60] Y. Ouyang, Z. Li, L. Su, W. Lu, and Z. Lin, Application behaviors driven self-organizing network (SON) for 4G LTE networks, *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 1, pp. 3–14, 2020.

[61] Y. Ouyang, Z. Li, L. Su, W. Lu, and Z. Lin, APP-SON: Application characteristics-driven SON to optimize 4G/5G network performance and quality of experience, in *Proc. 2017 IEEE Int. Conf. on Big Data* (*Big Data*), Boston, MA, USA, 2017, pp. 1514–1523.

[62] M. Qin, Q. Yang, N. Cheng, J. Li, W. Wu, R. R. Rao, and X. Shen, Learning-aided multiple time-scale SON function coordination in ultra-dense small-cell networks, *IEEE Trans. Wirel. Commun.*, vol. 18, no. 4, pp. 2080–2092, 2019.

[63] A. Engels, M. Reyer, X. Xu, R. Mathar, J. Zhang, and H. Zhuang, Autonomous self-optimization of coverage and capacity in LTE cellular networks, *IEEE Trans. Veh. Technol.*, vol. 62, no. 5, pp. 1989–2004, 2013.

[64] S. Berger, A. Fehske, P. Zanier, I. Viering, and G. Fettweis, Online antenna tilt-based capacity and coverage optimization, *IEEE Wirel. Commun. Lett.*, vol. 3, no. 4, pp. 437–440, 2014.

[65] R. Razavi, S. Klein, and H. Claussen, A fuzzy reinforcement learning approach for self-optimization of coverage in LTE networks, *Bell Labs Tech. J.*, vol. 15, no. 3, pp. 153–175, 2010.

[66] D. Karvounas, P. Vlacheas, A. Georgakopoulos, K. Tsagkaris, V. Stavroulaki, and P. Demestichas, An opportunistic approach for coverage and capacity optimization in Self-Organizing Networks, in *Proc. 2013 Future Network & Mobile Summit*, Lisboa, Portugal, 2013, pp. 1–10.

[67] R. Combes, Z. Altman, and E. Altman, Self-organization in wireless networks: A flow-level perspective, in *Proc. 2012 Proc. IEEE INFOCOM*, Orlando, FL, USA, 2012, pp. 2946–2950.

[68] L. Huang, Y. Zhou, J. Hu, X. Han, and J. Shi, Coverage optimization for femtocell clusters using modified particle swarm optimization, in Proc. *2012 IEEE Int. Conf. on Commun.*, Ottawa, Canada, 2012, pp. 611–615.

[69] B. Partov, D. J. Leith, and R. Razavi, Utility fair optimization of antenna tilt angles in LTE networks, *IEEE/ACM Trans. Netw.*, vol. 23, no. 1, pp. 175–185, 2015.

[70] D. Lee, S. Zhou, X. Zhong, Z. Niu, X. Zhou, and H. Zhang, Spatial modeling of the traffic density in cellular networks, *IEEE Wirel. Commun.*, vol. 21, no. 1, pp. 80–88, 2014.

[71] M. Amirijoo, L. Jorguseski, R. Litjens, and R. Nascimento, Effectiveness of cell outage compensation in LTE networks, in *Proc. 2011 IEEE Consumer Communications and Networking Conf.*, Las Vegas, NV, USA, 2011, pp. 642–647.

[72] A. J. Fehske, H. Klessig, J. Voigt, and G. P. Fettweis, Concurrent load-aware adjustment of user association and antenna tilts in self-organizing radio networks, *IEEE Trans. Veh. Technol.*, vol. 62, no. 5, pp. 1974–1988, 2013.

[73] S. Berger, M. Simsek, A. Fehske, P. Zanier, I. Viering, and G. Fettweis, Joint downlink and uplink tilt-based self-organization of coverage and capacity under sparse system knowledge, *IEEE Trans. Veh. Technol.*, vol. 65, no. 4, pp. 2259–2273, 2016.

[74] A. Awada, B. Wegmann, I. Viering, and A. Klein, A SON-based algorithm for the optimization of inter-RAT handover parameters, *IEEE Trans. Veh. Technol.*, vol. 62, no. 5, pp. 1906–1923, 2013.

[75] O. C. Iacoboaiea, B. Sayrac, S. B. Jemaa, and P. Bianchi, SON coordination in heterogeneous networks: A reinforcement learning framework, *IEEE Trans. Wirel. Commun.*, vol. 15, no. 9, pp. 5835–5847, 2016.

[76] J. Sánchez-González, J. Pérez-Romero, and O. Sallent, A rule-based solution search methodology for self-optimization in cellular networks, *IEEE Commun. Lett.*, vol. 18, no. 12, pp. 2189–2192, 2014.

[77] *Physical/electrical characteristics of hierarchical digital interfaces*, ITU-T G.703, 2016.

[78] *Network node interface for the synchronous digital hierarchy* (*SDH*), ITU-T G.707/Y.1322, 2007.

[79] Cisco ONS 15454 Series Multiservice Transport Platforms, https://www.cisco.com/c/en/us/support/optical-networking/ons-15454-series-multiservice-transport-platforms/series.html, 2013.

[80] *Spectral grids for WDM applications*: *CWDM wavelength grid*, ITU-T G.694.2, 2003.

[81] *Architecture for the automatically switched optical network* (*ASON*), ITU-T G.8080/Y.1304, 2001.

[82] NTT, CMCC, and ETRI, Packet transport networks: Overview and future direction, https://www.itu.int/en/ITU-T/C-I/interop/Documents/20140826/CI-2_INP-09_Packet_Transport_Networks_Overview_and_Future_Direction.pdf, 2014.

[83] P. Bhaumik, S. Zhang, P. Chowdhury, S. S. Lee, J. H. Lee, and B. Mukherjee, Software-defined optical networks (SDONs): A survey, *Photon. Netw. Commun.*, vol. 28, no. 1, pp. 4–18, 2014.

[84] R. W. Thomas, D. H. Friend, L. A. DaSilva, and A. B. MacKenzie, Cognitive networks, in *Cognitive Radio*, *Software Defined Radio*, *and Adaptive Wireless Systems*, H. Arslan, Ed. Dordrecht, Netherlands: Springer, 2007, pp. 17–41.

[85] A. Caballero, R. Borkowski, I. de Miguel, R. J. Durán, J. C. Aguado, N. Fernández, T. Jiménez, I. Rodríguez, D. Sánchez, R. M. Lorenzo, et al., heterogeneous and reconfigurable optical networks: The CHRON project, *J. Lightwave Technol.*, vol. 32, no. 13, pp. 2308–2323, 2014.

[86] I. T. Monroy, D. Zibar, N. G. Gonzalez, and R. Borkowski, Cognitive heterogeneous reconfigurable optical networks (CHRON): Enabling technologies and techniques, in *Proc. 2011 13th Int. Conf. Transparent Optical Networks*, 2011, pp. 1–4.

[87] J. Mata, I. de Miguel, R. J. Duran, N. Merayo, S. K. Singh, A. Jukan, and M. Chamania, Artificial intelligence (AI) methods in optical networks: A comprehensive survey, *Opt. Switch. Netw.*, vol. 28, pp. 43–57, 2018.

[88] C. Wang, R. Gu, Y. Xiao, W. Wang, X. Wang, and Y. Liu, IPv6+ based intelligent IP network solution, *Telecommun. Sci.*, (in Chinese), vol. 36, no. 8, pp. 66–80.

[89] IETF, SRv6 Network Programming: Draft-Ietf-Spring-Srv6-Networkprogramming-16, https://datatracker.ietf.org/doc/html/draft-ietf-spring-srv6-network-programming-16, 2020.

[90] Ö. Bulakci and Q. Wei, 5G mobile network architecture for diverse

services, use cases, and applications in 5G and beyond, https://5g-monarch.eu/wp-content/uploads/2019/05/5G-MoNArch_761445_D2.3_Final_overall_architecture_v1.0.pdf, 2019.

[91] *Study on enhancement of management data analytics*, 3GPP TR 28.809, 2019.

[92] L. Gupta, T. Salman, M. Zolanvari, A. Erbad, and R. Jain, Fault and performance management in multi-cloud virtual network services using AI: A tutorial and a case study, *Comput. Netw.*, vol. 1065, p. 106950, 2019.

[93] Mapping between ENI architecture and operational systems, https://portal.etsi.org/webapp/WorkProgram/Report_WorkItem.asp?WKI_ID=58945, 2020.

[94] *Experiential networked intelligence* (*ENI*); *Definition of data processing mechanisms*, ETSI GR ENI 009, 2021.

[95] *Experiential networked intelligence* (*ENI*); *InTent aware network autonomicity* (*ITANA*), ETSI GR ENI 008, 2021.

[96] ETSI, Ongoing PoCs, https://eniwiki.etsi.org/index.php?title=Ongoing_PoCs, 2020.

[97] ETSI, Elastic Network Slice Management, https://eniwiki.etsi.org/index.php?title=PoC_02:_Elastic_Network_Slice_Management, 2019.

[98] ETSI, Intent-based User Experience Optimization, https://eniwiki.etsi.org/index.php?title=PoC_08:_Intent-based_User_Experience_Optimization, 2020.

[99] Operations support system, https://en.wikipedia.org/wiki/Operations_support_system, 2021.

[100] Future OSS: Towards an open digital architecture, https://go.oracle.com/LP=87451, 2019.

[101] ECOMP (Enhanced Control, Orchestration, Management & Policy) Architecture White Paper, https://about.att.com/content/dam/snrdocs/ecomp.pdf, 2016.

[102] P. D. Bartoli, Defining the ECOMP architecture, https://about.att.com/innovationblog/041116ecomparchitect, 2016.

[103] GB921 business process framework (eTOM) suite, https://www.tmforum.org/resources/suite/gb921-business-process-framework-etom-suite/, 2020.

[104] The connecting with customers report. A global in-depth study of the online customer experience, https://docplayer.net/11270893-The-connecting-with-customers-report-a-global-in-depth-study-of-the-online-customer-experience.html, 2013.

[105] S. Magids, A. Zorfas, and D. Leemon, The new science of customer emotions, *Harv. Bus. Rev.*, vol. 76, no. 11, pp. 66–74, 2015.

[106] AsiaInfo Technology, AISWare CEM customer experience management, https://www.asiainfo.com/en_us/product_aisware_5g_cem_detail.html, 2020.

[107] F. F. Reichheld, The one number you need to grow, *Harv. Bus. Rev.*, vol. 81, no. 12, pp. 46–54, 2003.

[108] C. Stahlkopf, Where net promoter score goes wrong, *Harv. Bus. Rev.*, vol. 10, 2019.

[109] *Policy and charging control architecture*, 3GPP TS 23.203, 2005.

[110] *Experiential Networked Intelligence* (*ENI*); *ENI Definition of Categories for AI Application to Networks*, ETSI GR ENI 007, 2019.

[111] TM Forum, Autonomous networks: Empowering digital transformation for the telecoms industry, https://www.tmforum.org/resources/standard/autonomous-networks-empowering-digital-transformation-telecoms-industry/, 2019.

[112] H. Davies, What is 6G: Nokia will lead EU research with project Hexa-X, https://www.trustedreviews.com/explainer/what-is-6g-project-hexa-x-4112862, 2020.

[113] Next G Alliance, Building the foundation for North American leadership in 6G and beyond, https://nextgalliance.org/wp-content/uploads/2020/11/Next-G-Alliance-Founding-Members-Presentation-11162020.pdf, 2021.

[114] Focus group on technologies for network 2030, https://www.itu.int/en/ITU-T/focusgroups/net2030/Pages/default.aspx, 2020.

[115] ITU-T, A Blueprint of Technology, Applications and Market Drivers Towards the Year 2030 and Beyond, White Paper, https://www.itu.int/en/ITU-T/focusgroups/net2030/Documents/White_Paper.pdf, 2020.

[116] SA Plenary, https://www.3gpp.org/specifications-groups/28-specifications-groups/SA, 2021.

[117] 3GPP, Study on RAN-Centric Data Collection and Study on RANcentric Data Collection and Utilization for LTE and NR, https://www.3gpp.org/ftp/tsg_ran/TSG_RAN/TSGR_81/Info_for_workplan/revised_SID_11/RAN3_2, 2018.

[118] O-RAN use cases and deployment scenarios, https://assets-global.website-files.com/60b4ffd4ca081979751b5ed2/60e5aff9fc5c8d496515d7fe_O-RAN%2BUse%2BCases%2Band%2BDeployment%2BScenarios%2BWhitepaper%2BFebruary%2B2020.pdf , 2020.

[119] *Study on enablers for network automation for the 5G system* (*5GS*); *Phase 2*, 3GPP TR 23.700-91, 2019.

[120] M. Varela, P. Zwickl, P. Reichl, M. Xie, and H. Schulzrinne, From service level agreements (SLA) to experience level agreements (ELA): The challenges of selling QoE to the user, in *Proc. 2015 IEEE Int. Conf. on Communication Workshop*, London, UK, 2015, pp. 1741–1746.

[121] *Telecommunication management*; *Study on scenarios for intent driven management services for mobile networks*, 3GPP TR 28.812, 2018.

[122] *Network functions virtualisation* (*NFV*); *Management and orchestration*; *Report on management and orchestration framework*, ETSI GR NFV-MAN 001, 2021.

[123] D. M. Gutierrez-Estevez, M. Gramaglia, A. De Domenico, G. Dandachi, S. Khatibi, D. Tsolkas, I. Balan, A. Garcia-Saavedra, U. Elzur, and Y. Wang, Artificial intelligence for elastic management and orchestration of 5G networks, *IEEE Wirel. Commun.*, vol. 26, no. 5, pp. 134–141, 2019.

[124] *Definitions of terms related to quality of service*, ITU-T E.800, 2008.

[125] *Information technology — open systems interconnection — basic reference model*: *The basic model*, ISO/IEC 7498-1:1994, 1994.

[126] J. Crowcroft, S. Hand, R. Mortier, T. Roscoe, and A. Warfield, QoS' s downfall: At the bottom, or not at all! in *Proc. ACM SIGCOMM Workshop on Revisiting IP QoS*: *What Have we Learned*, *Why do We Care?*, Karlsruhe, Germany, 2003, pp. 109–114.

[127] M. Varela, L. Skorin-Kapov, and T. Ebrahimi, Quality of service versus quality of experience, in *Quality of Experience*, S. Möller and A. Raake, Eds. Cham, Germany: Springer, 2014, pp. 85–96.

[128] Q. Yang, Y. Liu, T. Chen, and Y. Tong, Federated machine learning: Concept and applications, *ACM Trans. Intell. Syst. Technol.*, vol. 10, no. 2, p. 12, 2019.

[129] S. J. Pan and Q. Yang, A survey on transfer learning, *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, 2010.