AI-Cube

oppo

6G AI-Cube Intelligent Networking

oppo

01 Current state of Functional Dimensions in 5G System

In order to enable a more effective and efficient network operation for the emerging mobile Internet services and the Internet of things services, the system design approach for the separation of control plane and user plane were adopted by 3GPP to to enable operators to fast track the implementation of service innovation to provide swift online and on-demand service for the product deployment. The user plane and control plane in 5G system are referred to, user-plane functions and control-plane functions, respectively.

The user plane corresponds to the performance dimension. It includes various user plane functions from the physical layer to the network layer to provide the communication ability and to enforce user data transmission performance in order to meet the specified data transmission requirements, such as delay, reliability, rate, jitter.

The control plane corresponds to the versatile dimension, which enables network functions to respond to the changing conditions (such as terminal movement, terminal and network load changes, etc.) in order to minimize and if possible, avoid any impact towards the normal data transmission. The versatile aspect also implies the level of responsiveness to react to subjective decisions. A network with swift response can quickly and accurately perform the system adjustments according to the subjective intentions.

Figure 1: Comparable of CP and UP with Versatile Dimension and Performance Dimension

As shown in the above Figure 1, the two logical planes are corresponding to the Performance and Versatile Dimension which represent all the necessary functionalities and capabilities provided by the network functions in the current 5G system. All KPI indicators defined by 5G [1] (Peak rate, user experience, spectrum efficiency, mobility control, delay, connection density, network energy consumption, traffic density, etc.) all are based on these two planes.

02 Intelligence dimension in 6G system

With the AI technology [2] evolves and becomes more sophisticated, and better understandings of use cases that benefit the AI-enabled network operation and application services become more apparent [3] The AI-enabled mobile network functions deliver the following system benefits.

Value Propositions from AI empowered mobile system

Value Proposition-1 **AI's assisted accurate decision-making ability**	Learning samples from the data collection with the assistance of AI reasoning provide accurate summary of the general rules between inputs and outputs to support decision making. When network functions faced with more and more complex scenarios, it is difficult to make quick and accurate decision relying on human experience. Using the outcome of the AI learning and training to match specific functional operation in different scenarios can provide more effective assistance on the decision making.
Value Proposition-2 **AI's powerful reasoning ability**	With the continuous improvement of network computing processing power, AI reasoning capabilities have become more effectual, which amplifies the usefulness provided by the AI algorithms and functions.
Value Proposition-3 **Self-evolving capabilities**	AI's reasoning functionality is not static due to the nature of the iterative operation in analytic learning and training. As the system environment could be changed over time, the analytic operation will continue to optimize and to refine the AI reasoning operation which assures the accuracy of the analytic outcome.
Value Proposition-4 **AI's transfer learning ability**	The trained AI model can be refined through learning and make adjustment corresponding to the changing conditions. In time, AI model becomes more sophisticated and it can apply to broader scope of system functions to deal with more variety of scenarios, providing the possibility to populate the AI and ML analytic results to more diversified applications.

The service/application that leverages AI technology in mobile system will go through a three-stage process: For-the-AI, By-the-AI, and Of-the-AI. For-the-AI refers to stage-1 support where the -mobile system is used as a transmission pipeline to support the application layer AI business, By-the-AI refers to the stage-2 support where part of the network functions (such as eMBB and URLLC) within the mobile system with some enhancements to support AI/ML operation based on architecture framework, Of-the-AI refers to the stage-3 support where the AI/ML functionalities as a fundamental system component to be imbedded into the mobile system architecture in order to achieve the objective of "Intelligence Everywhere".

5G system did not envision the AI empowerment at the beginning of the architecture design, the 5G/B5G system is limited to provide the For-the-AI and By-the-AI support because it did not consider how to support an effective intra-system communication among the network functions with built-in support for "intelligence dimension". As the mobile system is evolving towards 6G, it is a good opportunity to revamp the 6G mobile system architecture to incorporate the AI dimension so that the objective of supporting AI-integrated mobile system end-to-end – i.e. Of-the-AI, can be achieved. As the 6G AI-integrated mobile system becomes ubiquitous, one can foresee the accessible "Intelligence Internet" everywhere becomes a reality.

With the consideration to imbed AI functionality into every system components within the emerging 6G mobile system, a 3-dimentional functional plane is proposed as the pillar to construct the 6G mobile system architecture framework. As described earlier, the first is the Versatile dimension which is corresponding to the control plane; and the second is the Performance dimension which is corresponding to the user plane. The third is the Intelligence dimension which is newly introduced and is corresponding to the AI plane that provides the AI/ML data processing and analytic functionality. The benefits and the functionality provided by the AI dimension are very different from the existing operation on the CP/UP.

Features: The data transmission is content-oriented, and it is no longer purely transmitted between the terminal and the gateway through the pipeline. The data corresponding to the AI dimension will flow through different network functions when an update is needed. Furthermore, more than one network function could be updated in order to achieve the final reasoning and learning.

Characteristics: The AI data itself is valuable to the 6G network. The amount and the type of collected data directly influences the competency of AI capabilities. Therefore, in some cases, the collection of AI-related data will not be charged, instead, it will be compensated by the data provider.

Role: Data is used for training and reasoning, not just for meeting specific business needs or service obligation.

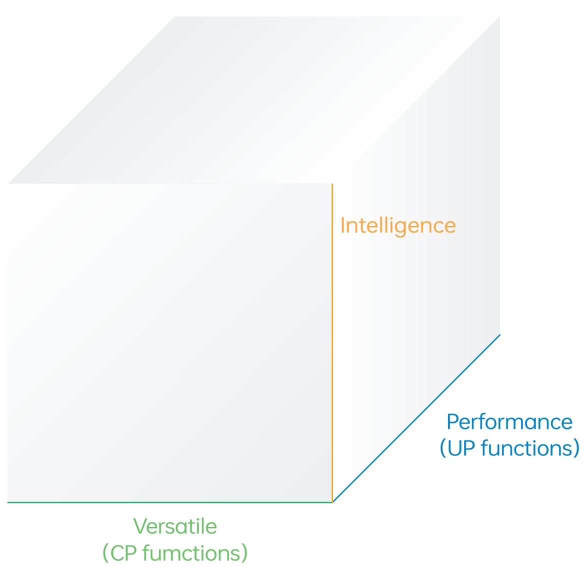

Figure 2: 6G space

Figure 2 above presents the introduction of new AI dimension in the 6G system, the 6G system can be visualized as a 6G space which is designed based on the three-dimensional forte as described above. The 6G system is composed of functional components of which the functionalities are corresponding to the three dimensions of the 6G space.

03 AI function plane and its empowerment of mobile network functions

3.1 AI functional plane

The 6G system will be implemented with the AI functionality built-in. The purpose of the AI function is to provide the necessary capabilities (e.g. AI models, analytical support, pre-processing data etc.) for the CP function and the UP function to enable AI functionality in 6G system. For this reason, the AI functional plane needs to support data collection/storage, machine learning and training, AI reasoning and model management/distribution [4]. These general AI functions are necessary means to empower the CP and the UP to support intelligent data analytics. In additional to the AI functions, new supporting functions are needed to provide additional capabilities specifically for the AI functional plane which are: AI connectivity support, AI domain management, and AI capability exposure as shown in Figure 3 below.

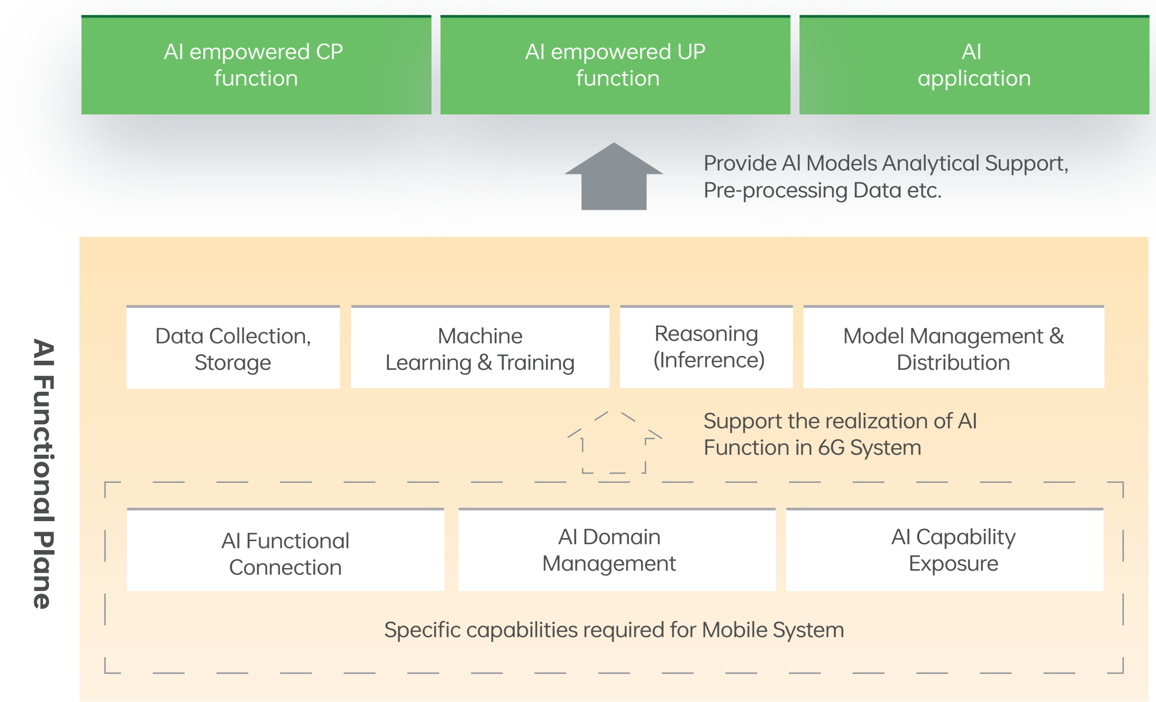

Figure 3: AI Functional Plane

The following focuses on the specific functions required by AI applications under mobile system, namely, AI functional connection, AI domain management, and AI capability exposure.

3.1.1 AI functional connection

AI data transmission and processing in the AI functional plane are the basis for all AI functions. Compared with the control plane (such as SRB, NAS connection) and user plane connections (such as DRB, GTP tunnel) in the 5G system, AI functional connection has the following uniqueness:

1)The AI data transmission path is determined on demand, and mass AI data transmission can be recognized and/or pre-processed between any two or more functional components in the 6G system through this connection. AI data transmission can be routed based on content and/or operational requirement, from the source functional component to one or more functional component towards the destination functional component;

2)Some or all of the functional components passing by the AI data transmission path can perceive the content and perform corresponding processing or pre-processing.

Figure 4 below describes an example for the analytics data processing and analytics data sharing along the AI connection. Functional components 1, 2, and 3 all store terminal user data, and perform big data analytic using federation mechanism. In each inference operation, the terminal could send the pre-processed data to the core network data analytics network function via the AI data connection over the AI functional plane. The core network data analytics network function further performs the pre-processing on the result provided by the terminal and also on the network data. The processed analytic data will then be sent to the edge server for use. In this process, the connection of the AI functional plane needs to pass through functional components 1, 2, and 3. All three components need to recognize and process related data.

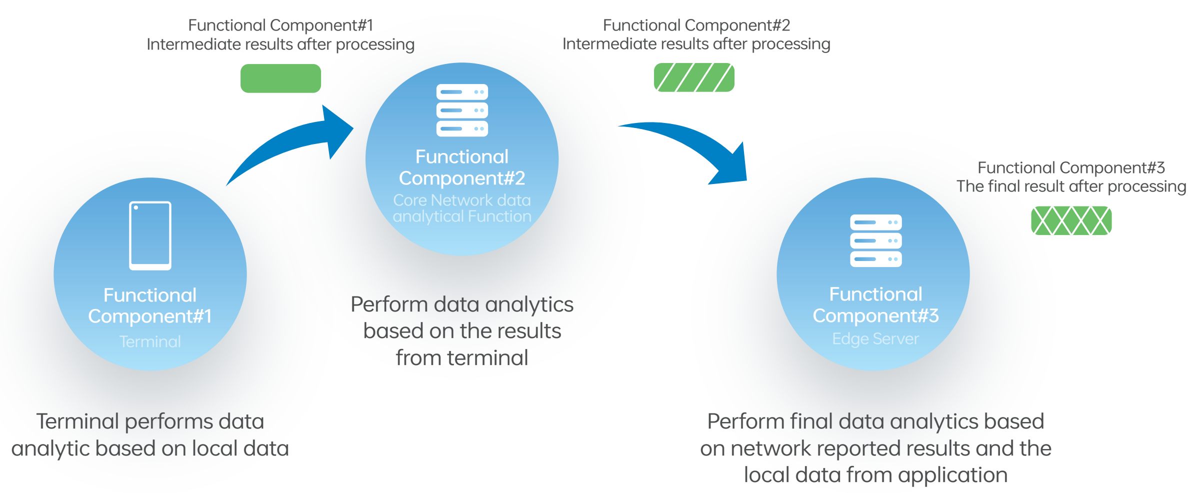

Figure 4: Operational flow between Functional Components on recognizing local analytics data processing and the analytic data sharing along the AI connection

3.1.2 AI Domain Management

For an AI model training or inference, the execution behavior needs to consider the following constraints:
Data restriction: As the collected data could implicate someone' s privacy, validity, etc., and the scope of application on analytic data is limited
Computing power restriction: the computing power usage allowed for a certain AI service is being controlled
Application condition restriction: AI services are limited by time, location, network status, and user behavior characteristics
Model restriction: Each model has its own limitation when applying to a certain kinds of scenarios

Therefore, AI operation in 6G system is often restricted and exclusive. In order to accurately define the scope of AI operation, we introduce the concept of AI domain. A 6G system can contain a variety of different AI domains, and each AI domain contains specific task, and AI domain resources for particular AI service. An AI domain is generally composed of multiple functional components in the network and variety of resources (computing, storage, data, communication resources, etc.) to perform AI model allocation and network resource scheduling for supporting the AI services , Data Collection and other needs:

AI domain resources, including computing, storage, data, communication, model library resources, etc., are generally only accessible by all functional components within the AI domain.

The AI domain performs overall management of the resources of each functional component - including the unified scheduling of data, AI models, computing capabilities, and communication capabilities. It makes full use of its own proprietary resources to optimize the AI model continuously.

Different AI domains can be divided by services, and can be further subdivided by location, user, third-party customization and other dimensions.

Figure 5: AI Domain Management

In 6G system, different AI domains isolate and restrict their corresponding resources to meet the needs of different AI operations. For terminals and users, joining a specific AI domain according to business requirements can obtain the most suitable AI model provided by the AI domain and required resources such as computing and communication support.

As shown in the example below in Figure 6, if a terminal needs to drive autonomously in a certain area, it needs to join the AI domain under that area, so that the AI domain can issue the most suitable AI model and work tasks for the terminal based on the characteristics and objective conditions of the terminal, and Allocate the most suitable edge server for it and establish the required communication connection. At the same time, the terminal can also participate in model training based on federated learning under the organization of the AI domain, which helps the AI domain of autonomous driving to further optimize the AI model.

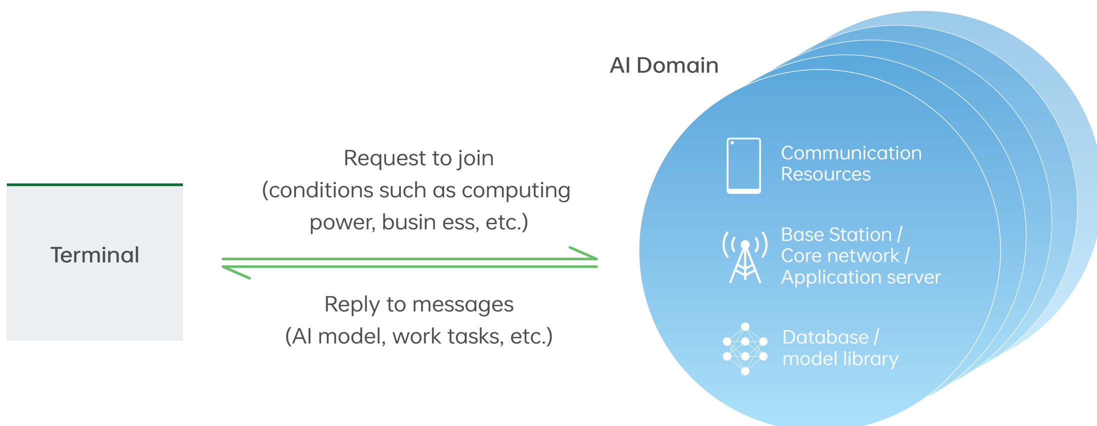

Figure 6: Example of terminal integration in AI domain operation

3.1.3 Capability exposure in AI functional plane

Capability Exposure is an important function of mobile networks, and it already supported in 5G system [5]. In the 6G era, the exposure of AI-related capabilities is an important part of the open interface strategy for 6G system capabilities and an important entry point for the integration of mobile network with ICT (Internet and Communication Technology). The capability exposure of AI functions includes the following three aspects:

1) Data exposure: One of the important factors for the large-scale globalization of AI technology is data exposure [6]. To train a good model and to develop robust reasoning capabilities, cross-domain data collection is a necessity. In the context of this paper, the main focus is the three different domains: the terminal, the mobile system and the OTT which share similar level of security and privacy. Just like the human brain to make decision, only when more comprehensive information is provided, the more reasonable the judgments can be made. Therefore, the exposure data is a key factor that determines whether the AI functionality in the 6G system is advance or not.

2) Resource exposure: Many AI-based services need to consume a lot of computing, storage, and communication resources [7], and most OTT (Over The Top) manufacturers do not have such a huge amount of resources. The operator's network can provide cloud resource services for OTT, allowing OTT to use the operator's cloud resources to perform corresponding AI services or to perform AI services for OTT companies. This is also a good opportunity to truly realize the integration of ICT at the business level.

3) Model exposure: AI models often have the learning transfer capabilities. AI models can be shared between 6G system and OTT, and pre-trained models can be obtained on demand [8] to create a new kind of business, e.g. "AI model application stores" .

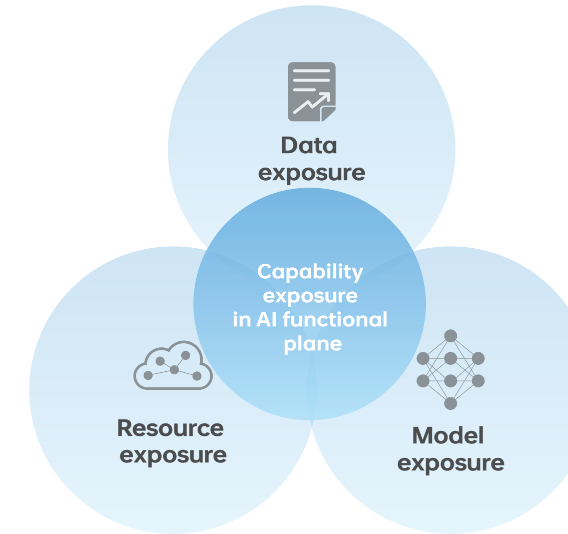

Figure 7: Capability exposure in AI functional plane

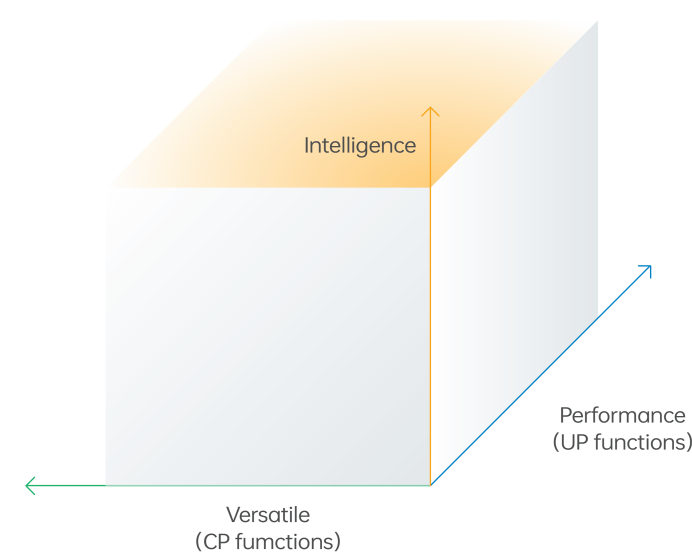

Figure 8: AI Functional Plane in 6G Space

Based on the descriptions above, one would recognize the AI functional plane supports AI-related functions through the unique features of the 6G system (AI functional plane connection, AI domain management, AI-related capability exposure), thereby providing AI-enabled 6G system capabilities (Models, Data, Analytics, etc.) to enable AI empowerment of CP and UP functions. Therefore, the ability of the AI functional plane will determine the level of intelligence of the CP function and the UP function. As shown in the Figure 8 below, the AI functional plane determines the upper bound of the intelligence dimension of the 6G system in the 6G space. The more advance of the AI function surface ability, the higher level of intelligence that empowers CP/UP.

3.2 AI empowered 6G System

The AI functional plane empowers the traditional control plane functions and user plane functions in 6G system to form an AI-powered user plane and an AI-powered control plane respectively. AI empowerment includes two types of methods: decision-making and non-decision-making which are described further in the following Table 3.2-1. 6G system empowerment through AI can either expand or enhance its AI solutions based on existing functions, or it can overturn existing algorithms and logic and use AI model reasoning to completely replace them.

Table-3.2-1: Understanding Decision-making vs. Non-decision making AI-empowerment

	Characteristics	Usage	Self-evolution and self-optimization capabilities
Decision-making AI empowerment	AI reasoning is used to guide system behavior, and the reasoning result itself will not be transmitted to the target end. Small amount of data	Generally used for strategy formulation, mostly used for control plane related functions	Yes
Non-Decision-making AI empowerment	AI reasoning is directly used to process business data, and the reasoning result itself is transmitted to the target end as data. Large amount of data	For the processing of business original data, it is mostly applied to user-related functions	Yes

3.2.1 Decision-making AI empowerment

A common form of AI empowering mobile networks is to make more reasonable and flexible decision-making configurations for specific functions based on richer dimensions, which are referred as decision-making methods.

The following Figure 3.2.1-1 illustrates the method of AI empowering decision-making functions. As shown by the orange arrow, various functions can help to achieve intelligent improvement through two types of inputs.

The first category of input is the common prediction, including service prediction, location prediction, load prediction, and user behavior prediction. Each specific function uses one or several common predictions as an important basis for input judgment. In order to achieve common prediction, it is necessary to collect data from the terminal, network and application.

The second category of input is individual data. When AI enables different functions, there will be different individual input parameters. For example, RSRP and measurement event configuration, etc. are individual input parameters for handover function. Cell interference conditions and historical random access data, etc. are individual input parameters for random access function.

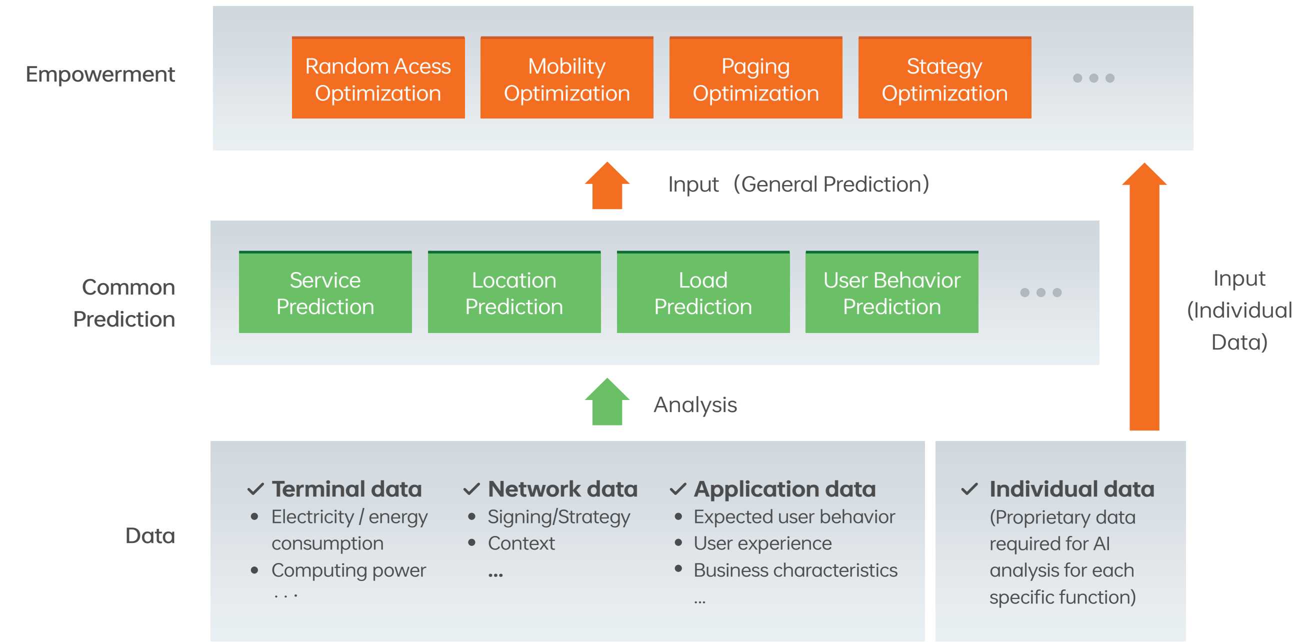

Figure 9: Decision-making AI Empowerment

In the actual AI empowerment, in addition to the need to define input parameters and output parameters. As shown in Figure 10 below. "Common prediction" and "Individual data" are the input content shown in Figure 9, and the individual output is the output result of the AI model analysis. It is necessary to define an individual output for each specific function point in order to achieve reasonable, fast and accurate decision-making (Chapter 3.4 gives specific examples of how to achieve AI-enabled switching by defining general predictions, individual data, and individual output).

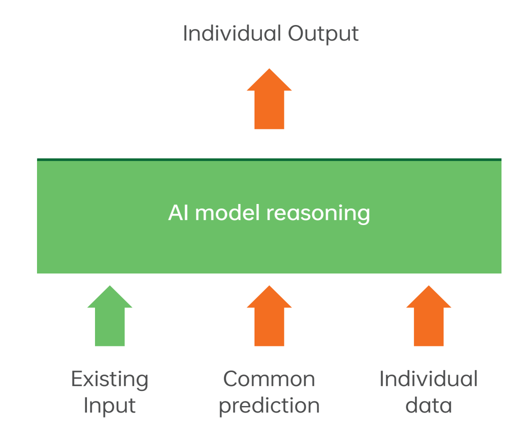

Figure 10: Decision-making AI Empowerment

3.2.2 Non-decision-making AI empowerment

The characteristics of non-decision-making AI empowerment is that it no longer uses the algorithms and processing steps that have been defined by 3GPP, and it may even departs from the current understanding of a certain function of the communication system, and uses AI reasoning instead. The output data generated by this type of AI empowerment is generally transmitted to the target as content. At present, with the development of artificial intelligence technology and the expansion of computing power resources, it has gradually become possible to use AI methods to optimize and solve traditional complex problems [9]. For example, under such a premise, AI model inference could be used to provide channel coding and decoding result in Radio Access Network (RAN)., In other words, AI model inference, could become a new approach to develop future system transmission schemes in AI-enabled 6G system via the AI process of construction, self-optimization and self-evolution.

As shown in the Figure 11, the specific form of non-decision-making functions empowered by AI is to replace part or all of the steps of data processing through AI model reasoning. For each function, the limiting factors and expected effects need to be defined individually. AI model can perform the best data processing according to the above two personalized definitions. Through AI empowerment, the data processing process that originally required human maintenance and the iterative optimization will now become self-learning and self-evolving operations empowered by AI. The ultimate goal of AI empowerment is to achieve continuous optimization in performance and accuracy (see section 3.4 for detailed examples).

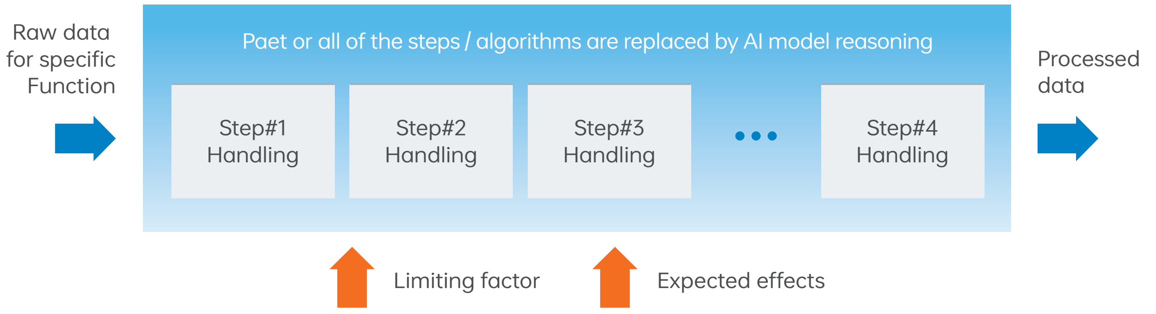

Figure 11: Non-decision-making AI empowerment

3.3 AI empowered Control Plane (AI powered UP)

Based on the AI empowerment methods introduced in the previous chapters, many legacy control plane functions can be empowered. As a result, the CP function can be refined in the vertical AI dimension, thereby forming an "AI empowered control plane" in the 6G space, as shown in the following Figure 12 in the green.

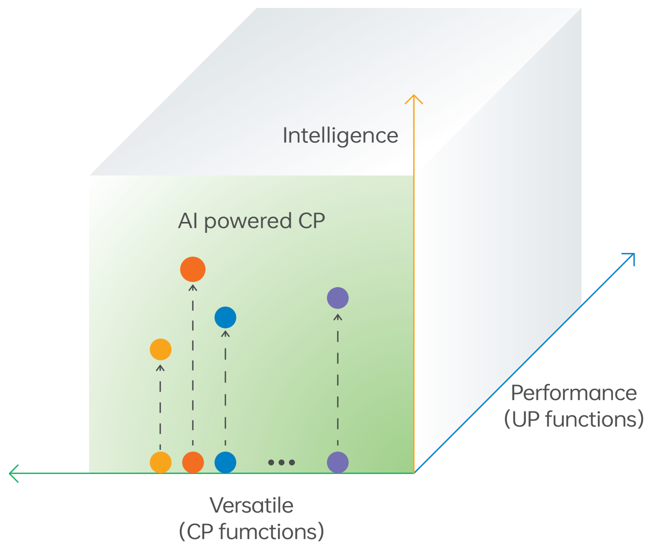

Figure 12: AI Empowered Control Plane

Since most of the control plane functions are to implement the formulation of strategies, most of the control plane functions are considered to be empowered by decision-making methods to achieve optimal effects. As described in section 3.2.1-1 previously, the AI empowerment of each CP function needs to define input parameters (including general predictive input and individual input parameters), and also needs to define individual output parameters.

Taking the handover function as an example, handover is one of the most distinctive control plane functions in mobile system. AI-based handover decision-making and service prediction can help increase the probability of handover success and reduce the impact on user experience.

As shown in the Figure 13 below, the legacy handover is based on the cell measurement results, and its output is only the decision to make a handover. The switching function based on AI empowerment is shown on the right side of the Figure 13 below, and its input parameters include

- Legacy input (cell measurement results)
- One or more general prediction (business prediction, location prediction, load predictions, user behavior prediction)
- Individual input parameters (current terminal speed, upper limit of delay, reliability requirements, etc.)

These three input parameters (legacy input, general prediction input, and individual data input) are used together as the input of the AI model and infer the individual output. The individual output here is a dedicated output customized to optimize the switching function. It not only includes the decision of whether to switch, but also the switching type (such as legacy HO, DAPS, CHO, RACH-less, etc.) and switching conditions (such as the target cell list), Time, location information, etc.) decision-making information, and a wealth of individual output parameters can control switching more flexibly, thereby improving the success rate of switching and ensuring service continuity during the switching process.

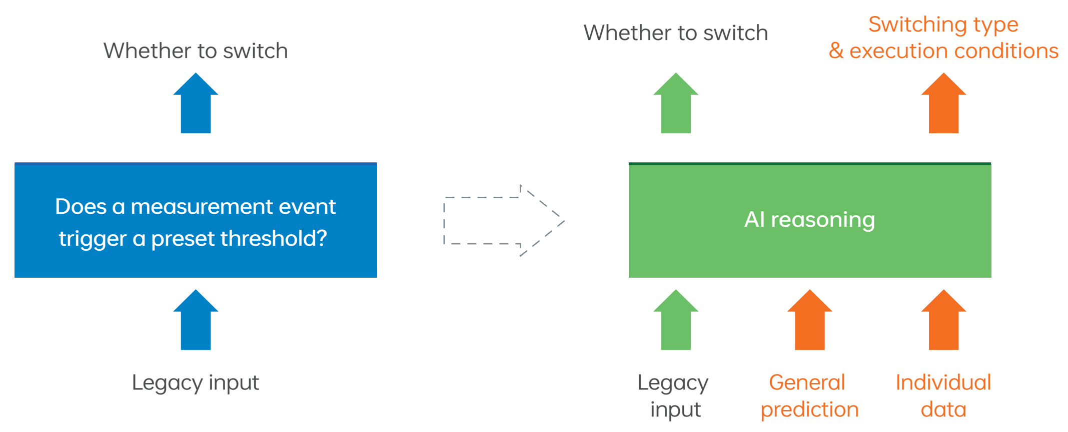

Figure 13: Example of decision-making AI Empowerment - The Switching Function

In addition to using the above AI empowerment method to enable the handover function, other control plane functions can also be improved using the above method, such as: AI empowered access control process、AI empowered terminal energy saving、AI empowered channel estimation, etc.

3.4 AI empowered User Plane (AI powered UP)

Another example is to leverage AI empowerment methods as introduced in the previous chapter to empower legacy user plane functions, the user plane functions could then be improved within the vertical AI dimension and to become an "AI-empowered user plane" in the 6G space, as shown in the blue part of Figure 14 below.

Since most of the user plane functions are to process the original data and to send the processed data to the opposite end to attain the required transmission requirements and user experience, most of the user plane functions can use non-decision-making methods for AI empowerment. It uses AI model reasoning to replace legacy algorithms and functional modules to achieve the same performance or even exceed the original performance. More importantly, it has the ability to be self-improved, to become self-evolution and self-optimization.

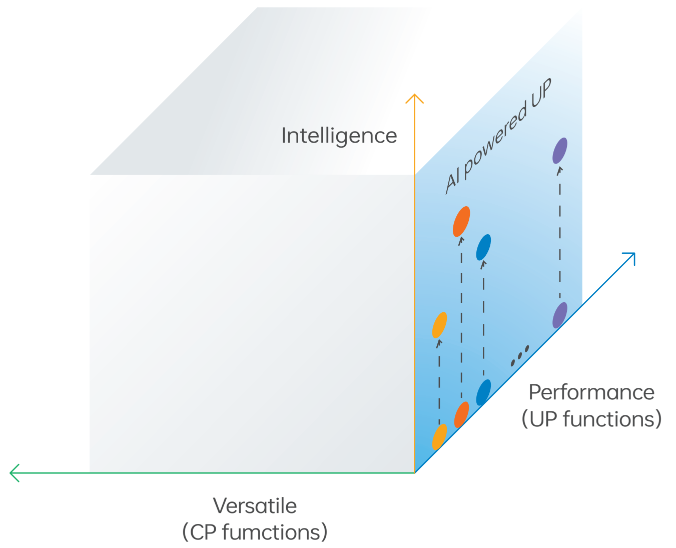

Figure 14: AI Empowered user plane

As shown in the Figure 15, the process of channel coding and decoding process is used as an example for using AI model inference to replace the existing wireless channel coding and decoding modules to obtain good signal gains.

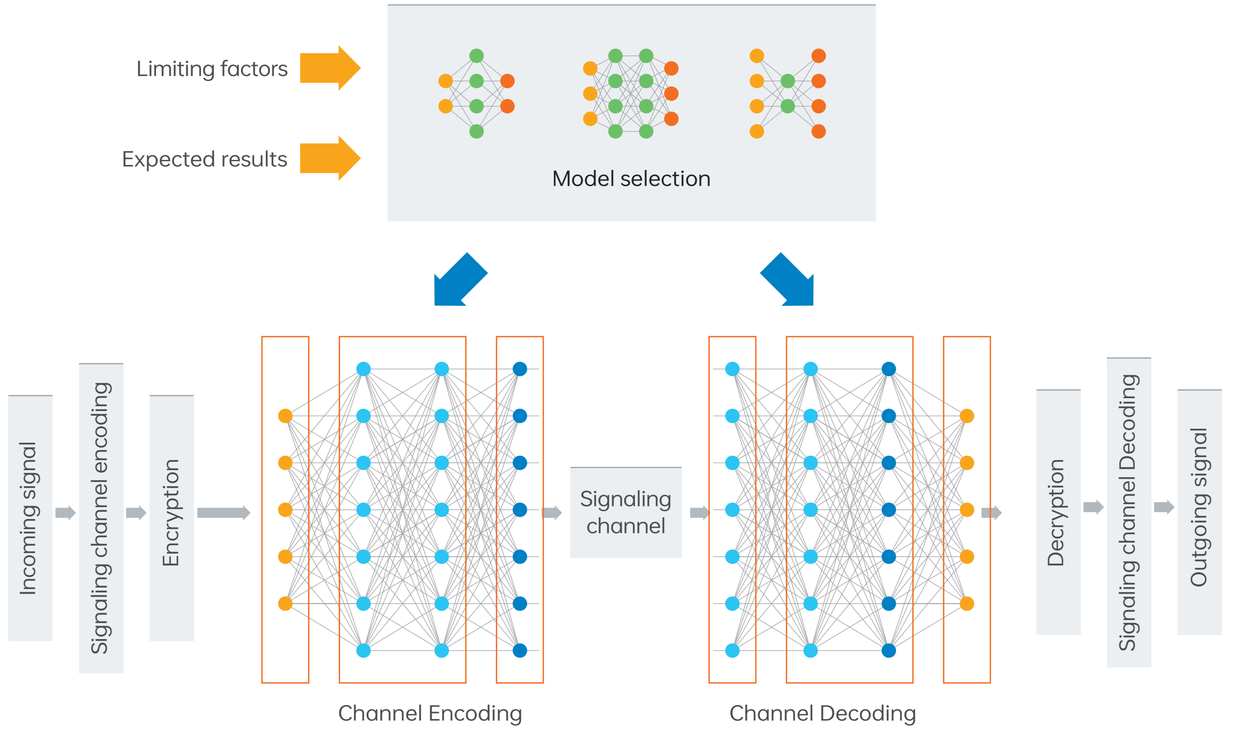

Figure 15: Use AI model for channel encoding and decoding

In the AI empowerment process, the limiting factors and expected results need to be defined on demand:

Limiting factors: including channel conditions, bandwidth limitations, delay sensitivity, burst errors, etc.;

Expected result: including transmission rate requirements, bit error rate requirements, reliability requirements, etc.

The two above inputs are used to select an appropriate AI model for channel coding to achieve the optimal effect in a specific environment and specific conditions, and to improve the performance of the entire communication system. The selected AI model replaces the existing channel codec module to achieve the AI empowerment effect. Through the self-learning and self-optimization of the AI model, the model can adapt to changing limiting factors and ensure the expected output, so as to reach or to exceed the signal gains that legacy channel coding brings to communication.

04 AI-Cube

4.1 Qualitative measurement of three dimensions in the 6G space

Before introducing AI-Cube, we first qualitatively describe the three dimensions in the 6G space as introduced in the previous chapter - i.e. how to qualitatively measure the performance dimension (known as the user plane function), the versatile dimension (known as the control plane function) and the size of the Artificial Intelligence dimension.

1) For the measurement of performance dimension

The measurement of the performance dimension depends on two aspects:

1. the communication capability - It includes the requirements for speed, reliability, delay, jitter, etc., and

2. the ability to provide application services - 6G system can participate in supporting application services by providing computing power and storage, rather than just being used as a pipeline for data transmission. Therefore, the performance of the 6G system can be comprehensively evaluated through two criteria: communication capability and application service capability.

2) For the measurement of versatile dimension.

The measurement of the versatile dimension depends on the network's ability to withstand changes in conditions. The conditions here include the user's moving speed, the execution speed of policy changes, the response speed of network status changes, and the precision of control. Based on the criteria above, the versatility of the 6G system can be comprehensively evaluated. Generally speaking, the greater of the versatility, the lesser the impact from the change of conditions on the performance, and the higher precision on the accuracy of network control.

3) For the measurement of Intelligence dimension

Intelligent of measurement is the most challenging part. The following two criteria are used to measure.

The first is the degree of substitution. This refers to the degree of AI capability substitution by using an AI model to replace a module or a parameter in the wireless communication capability, and may be gone even further, by having the entire system to be implemented end-to-end imbedded into AI operation. The score on the intelligence dimension is directly proportional to the degree of substitution of AI. The more substitution at the system level, the higher the score.

The second is from operational efficiency. This refers to the standardized AI operation behavior (such as a standardized AI model), and then calculate the number of AI operations completed under specific conditions (including but not limited to a certain time, frequency, computation capability, storage, energy consumption, etc.). The score on the intelligence dimension is directly proportional to the efficiency of its computing AI operation. The higher the AI operation efficiency, the higher the score.

In summary, the scoring of the intelligence dimension of a network is determined by the degree of substitution and operational efficiency.

4.2 AI-Cube

The 6G system can be qualitatively measured in the three dimensions of the 6G space. Therefore, the AI-powered user plane, the AI-powered control plane and the AI functional plane in the 6G space are bounded, and these three planes are jointly enclosed a three-dimensional space, which is referred as AI-Cube, as shown in Figure 16. The three planes in the AI-Cube jointly give new vitality to the 6G system.

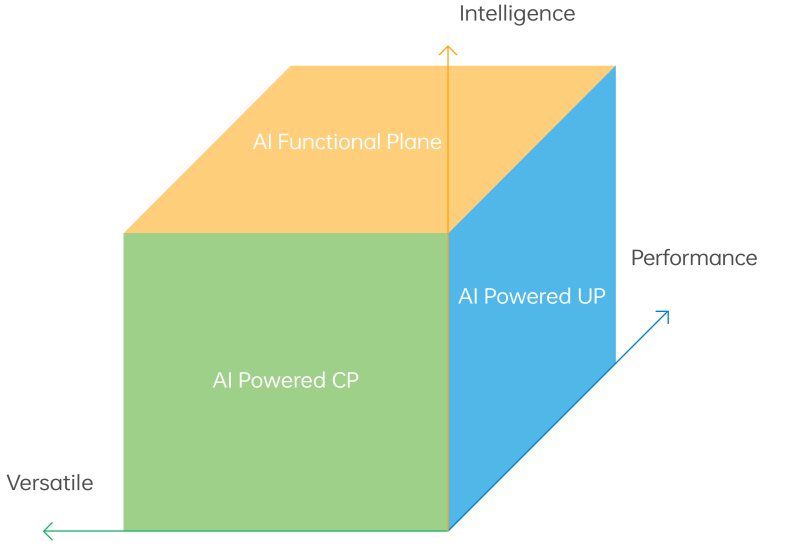

Figure 16: AI-Cube

The capabilities of the 6G system in the three dimensions 6G space determine the capacity and performance of this AI-Cube space. Through the AI functional plane of AI-Cube, rapid subnet establishment based on scenarios and requirements can be recognized, so that the appropriate 6G system functions can be selected accordingly, and 6G system can be partitioned into various individual subnets to serve diverse usage scenarios.

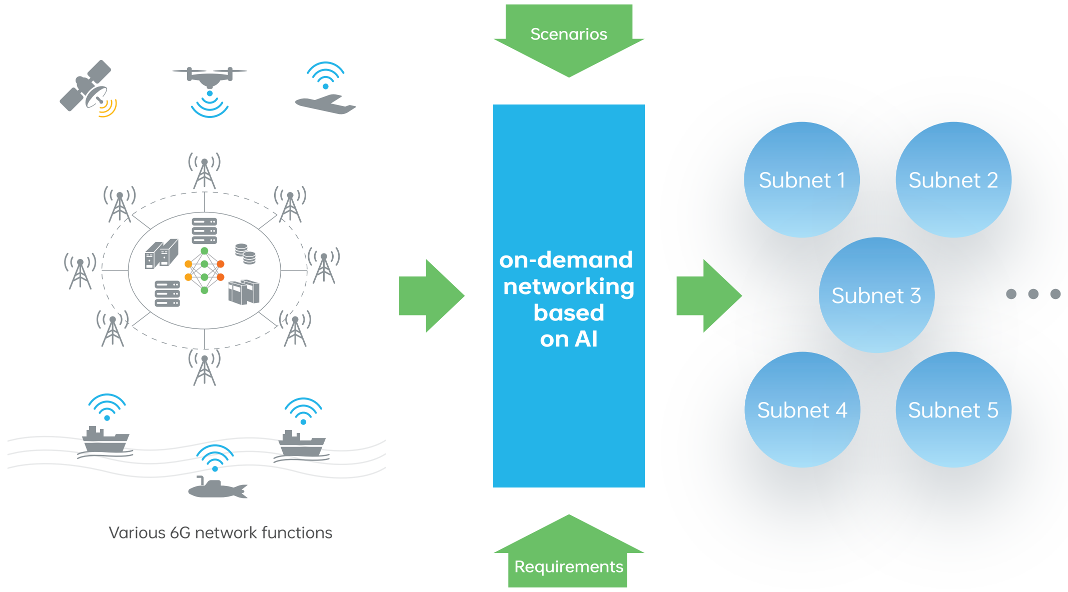

Figure 17: 6G System Capability for Subnets Partitioning

Different subnets partitioned of the 6G system can be identified by different area in the AI-Cube space according to their characteristics, as shown in Figure 17 for example.

Subnet 1: Smart factory (high performance): It has very important requirements for high reliability, small delay, and high speed to meet the precise production of the production line;

Subnet 2: Smart home (high intelligence, high flexibility): the user's mobile phones, bracelets, TVs, air conditioners and other devices are connected together and can communicate with each other;

Subnet 3: Intelligent driving (high intelligence, high flexibility), which requires low network transmission delays and powerful computing capabilities. In addition, vehicle trajectory prediction and risk assessment under specific scenarios also require high accuracy.

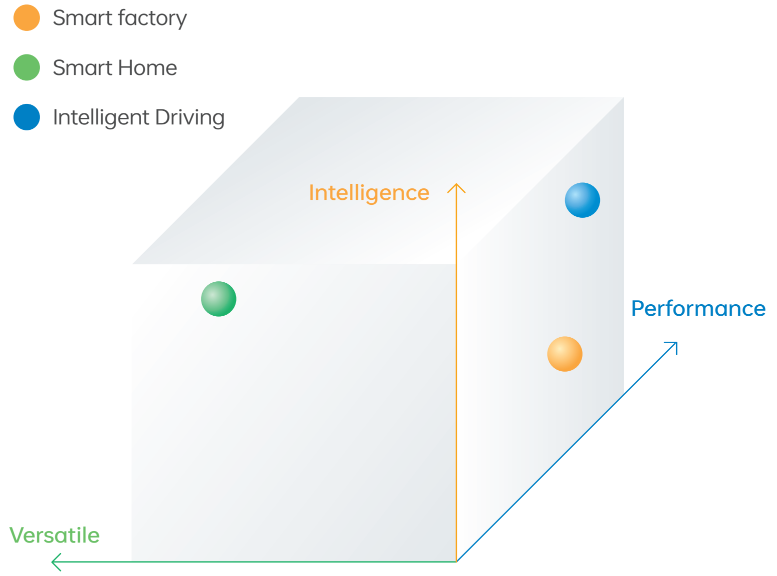

Figure 18 : Subnet in AI-Cube

05 AI Cube enabled 6G system architecture

The 6G system architecture empowered by AI-Cube is a mobile communication architecture based on infrastructure. Each system component and terminal in the network infrastructure can provide related resources (computing, storage, network resources, etc.) for each plane. The planes cooperate with each other and combine capability exposure to support data intercommunication with third-party applications, so as to enable the AI empowerment of 6G system.

As shown in Figure 19, the 6G system architecture includes an AI functional plane. The AI functional plane provides corresponding AI model configurations for various functions of the control plane and user plane on demand in order to enable various existing control plane functions and user plane functions to become the AI-powered user plane and AI-powered control plane. At the same time, the AI-enabled UP and CP send AI reasoning and other network state-related data to the AI functional plane as needed for AI self-optimization and self-evolution.

The three planes in the 6G system are coordinated by the joint allocation of resources between the terminal and the network side. As an important part of the 6G system, the terminal can provide valuable computing power and data resources to complete related AI functions.

The data that interacts with OTT applications through the capability exposure is an indispensable part of AI empowerment. OTT can exchange data with three planes:

The interactions between OTT applications and AI functional plane can be used for cross-domain data and model management and sharing to achieve more powerful AI functions.

The interactions between OTT applications and the AI-powered control plane can be used to obtain application information (such as data characteristics, traffic forecasts, etc.) on the plane to meet high versatility requirements. On the other hand, it can share data related to 6G control functions to help adjusting the behavior of the application layer.

The interaction between OTT applications and the AI-powered user plane can be used to perceive the business content of data packets in real time, and enable intelligent resource scheduling and routing based on business content to ensure the best data transmission rate, delay, and reliability.

The cross-domain cooperation ability between the three planes and OTT is the key for 6G system to support ICT integration. By recognizing the AI technology as an opportunity to enable the full system interactions and integration between terminal, network and OTT, we can then achieve a qualitative leap of 6G system functionality based on AI technology and fulfill the 6G vision of "Of-the-AI" .

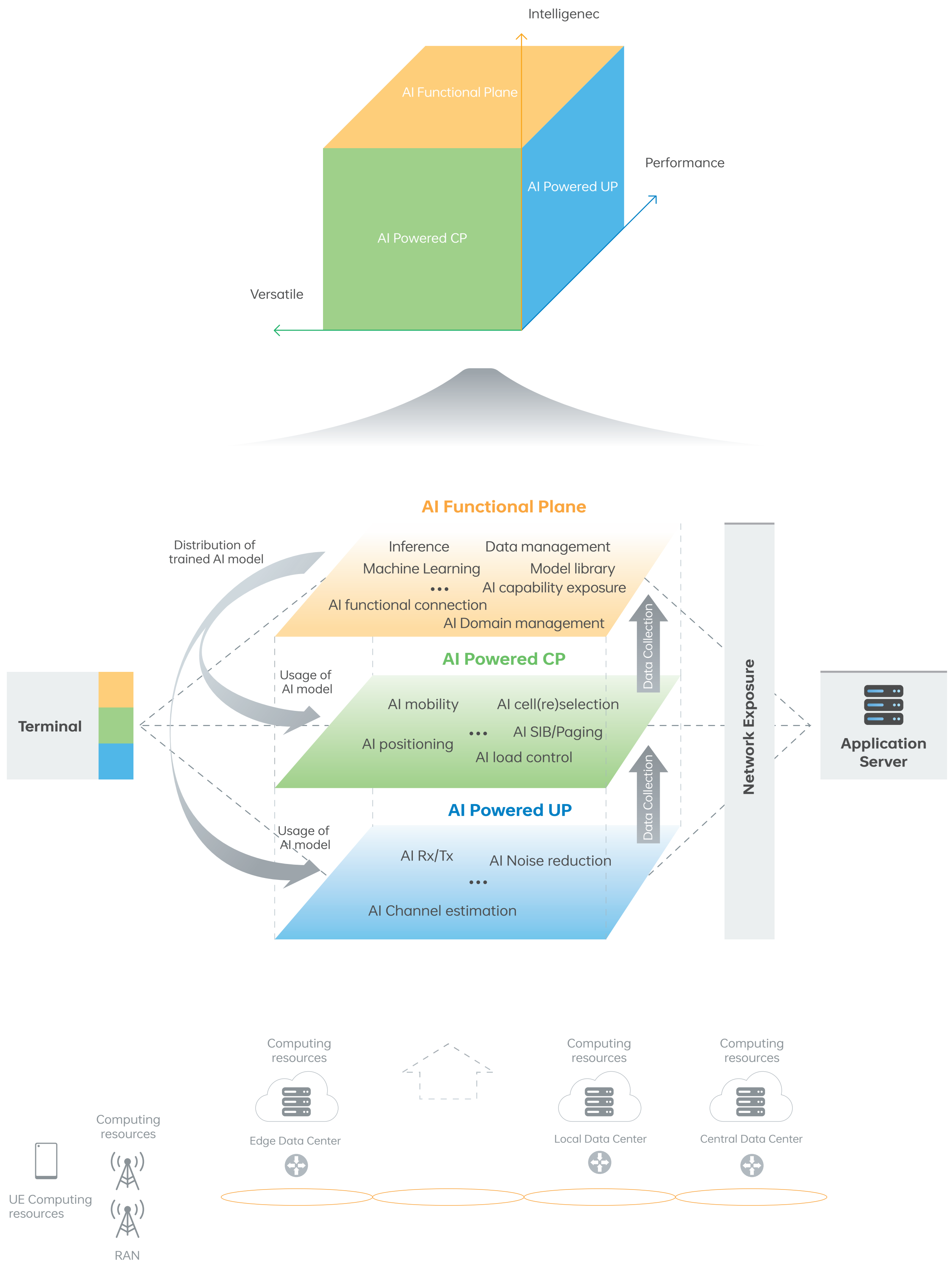

Figure 19: 6G network architecture empowered by AI-Cube

06 Looking Forward

The entire system integration with AI technology in mobile system is the most important feature of the 6G system. This paper describes the AI-enabled 6G system architecture framework to target for extreme optimization of the network control plane and user plane functions by integrating the AI plane functionality with the two legacy system planes to substitute the legacy functionality with AI-enabled functionality. This paper also describes to take the advantages of its huge resources of the mobile system to provide AI-enabled functionality support with OTT services. By enabling end-to-end AI-enabled functionality support with the terminals (i.e. UE domain) and OTT (i.e. Application domain), an "intelligence everywhere" becomes the ecosystem home for 6G.

The "three AI-enabled planes" (AI functional plane, AI-powered user plane, AI-powered control plane) and "AI capability exposure" (i.e. exposing AI related capability to enablecross-domain cooperation) based on concepts of AI-Cube are the key pillars to establish AI-enabled 6G ecosystem. By focusing on the principle of AI-Cube, an "endogenous intelligence" 6G network will be realized.

Reference

[1] ITU-R. ITU towards "IMT for 2020 and beyond" [R]. [S.l.:s.n.], 2015.

[2] Pouyanfar S, Sadiq S, Yan Y, et al. A survey on deep learning: Algorithms, techniques, and applications[J]. ACM Computing Surveys (CSUR), 2018, 51(5): 1-36.

[3] Chen M, Challita U, Saad W, et al. Artificial neural networks-based machine learning for wireless networks: A tutorial[J]. IEEE Communications Surveys & Tutorials, 2019, 21(4): 3039-3071.

[4] Miao, Hui, et al. "Towards unified data and lifecycle management for deep learning." 2017 IEEE 33rd International Conference on Data Engineering (ICDE). IEEE, 2017.

[5] GPP TS 23.501: "System Architecture for the 5G System; Stage 2" [S]. 3GPP, 2021.

[6] L' heureux, Alexandra, et al. "Machine learning with big data: Challenges and approaches." IEEE Access 5 (2017): 7776-7797.

[7] Wang, Xiaofei, et al. "In-edge ai: Intelligentizing mobile edge computing, caching and communication by federated learning." IEEE Network 33.5 (2019): 156-165.

[8] Niu, Shuteng, et al. "A decade survey of transfer learning (2010–2020)." IEEE Transactions on Artificial Intelligence 1.2 (2020): 151-166.

[9] Wang, Cheng-Xiang, et al. "Artificial intelligence enabled wireless networking for 5G and beyond: Recent advances and future challenges." IEEE Wireless Communications 27.1 (2020): 16-23.

Abbreviation

3GPP 3rd Generation Partnership Project
5G 5th Generation Mobile Communication Technology
UP User Plane
CP Control Plane
KPI Key Performance Indicator
AI Artificial Intelligence
B5G Beyond 5G
6G 6th Generation Mobile Communication Technology
OTT Over The Top
SRB Signalling Radio Bearer
NAS Non-Access Stratum
DRB Data Radio Bearer
GTP GPRS(General packet radio service) Tunneling Protocol GPRS
ICT Information and Communications Technology
RSRP Reference Signal Receiving Power
HO Handover
DAPS Dual Active Protocol Stack
CHO Conditional Handover
RACH Random Access Channel

AI-Cube