

OPPO View Towards SA1 6G

2024-5

Potential directions for 6G study

■ 6G Goal:

- to provide substantial value to users and industry by fostering new value-adding services
- to reduce total cost of ownership (TCO) of network and UE by adopting high-efficient, cost-effective, sustainable, resilient technologies

■ Proposed potential directions for 6G study:

eMBB

- XR/Metaverse
- User centric service

Computation + AI

- On-demand connection topology
- Computation network
- Data plane/service
- Model LCM
- Unified exposure for AlaaS

Sensing

- UE assisted sensing
- Multi-dimensional sensing
- Sensing for low altitude economy (e.g. UAV detection)

Vertical

- HRLLC
- TSN
- NPN

Ubiquitous connectivity

- NTN
- Ambient_IoT

Network operation

- Autonomous Networking
- Minimized Kernel + N subsystem
- Migration/Interworking
- Energy efficiency

Computation + AI

- 6G network can be deeply involved into a certain AI task, by means of providing QoAIS (Connection, Data, Computation, Model)
- AlaaS should be a critical capability exposure platform to support both MNO own services and 3rd party services

Connection

- Differentiated AI data transmission services
- Task-centric connectivity networking
- The content can be aware and processed on-path

Computation

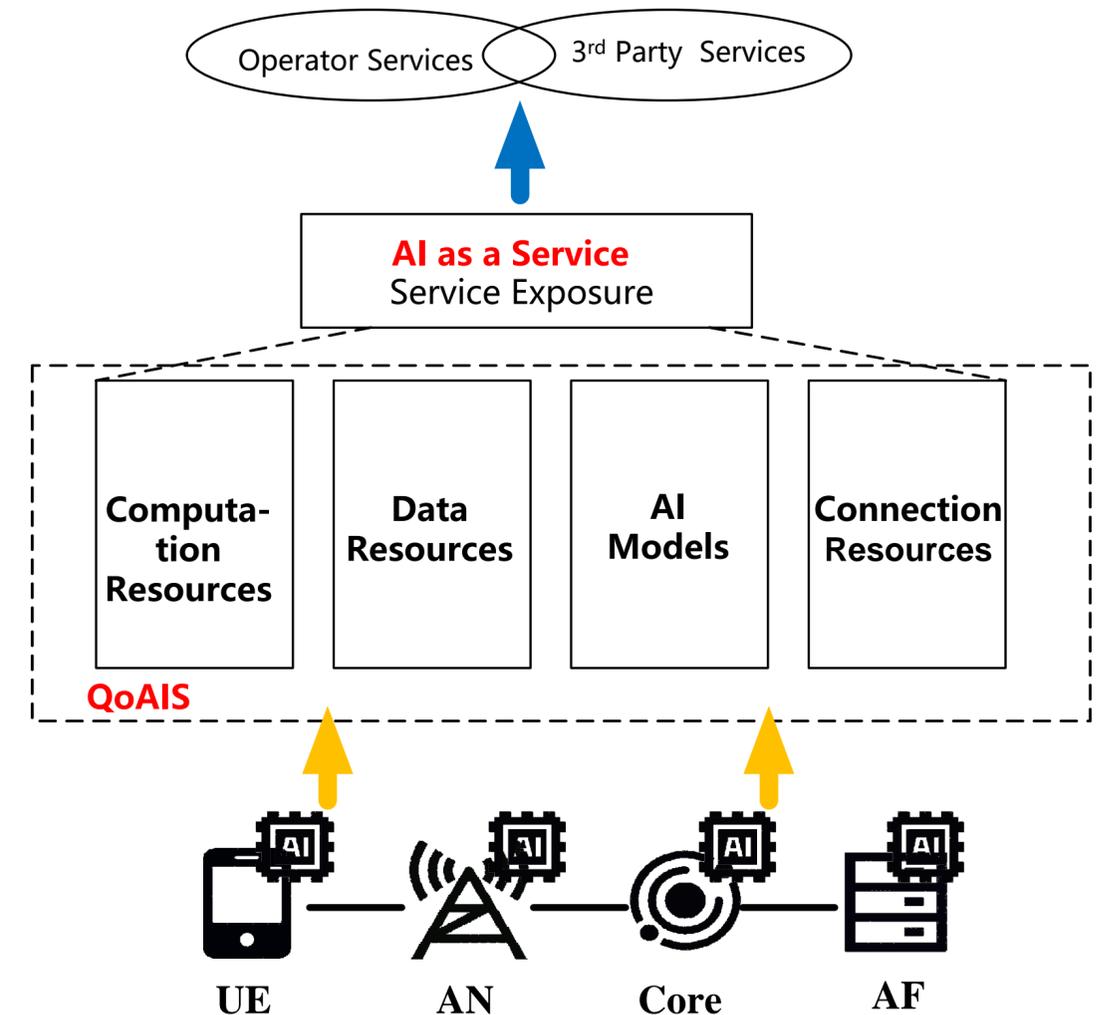
- Metric and scheduling of heterogeneous computation resource
- The trade-off between computation and communication resources
- KPI: Computation + Communication

Data

- Distributed /Cross-domain data
- Data ownership
- Realtime Data and information timeliness

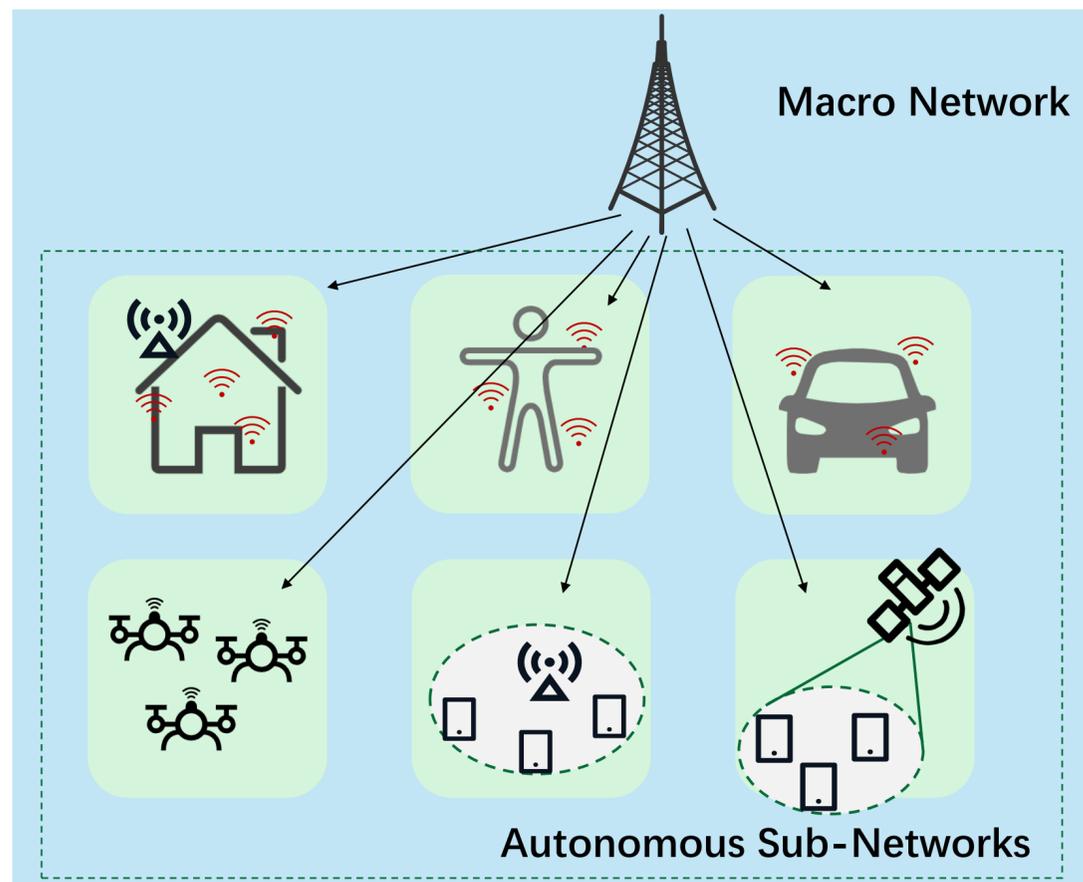
Model

- Life Cycle Management, incl.
 - Fine-tuning
 - Accuracy
 - Generalizability
 - ...
- Explain-ability/Trustworthy



Autonomous Networking

- 6G services have extreme QoS requirements, which is achievable / typical happens proximity, e.g., In-Vehicle, In-Body, In-Home, In-Robot.
- A plug-and-play networking concept: Autonomous Sub-Networks



Macro Network (i.e., Macro xNB)

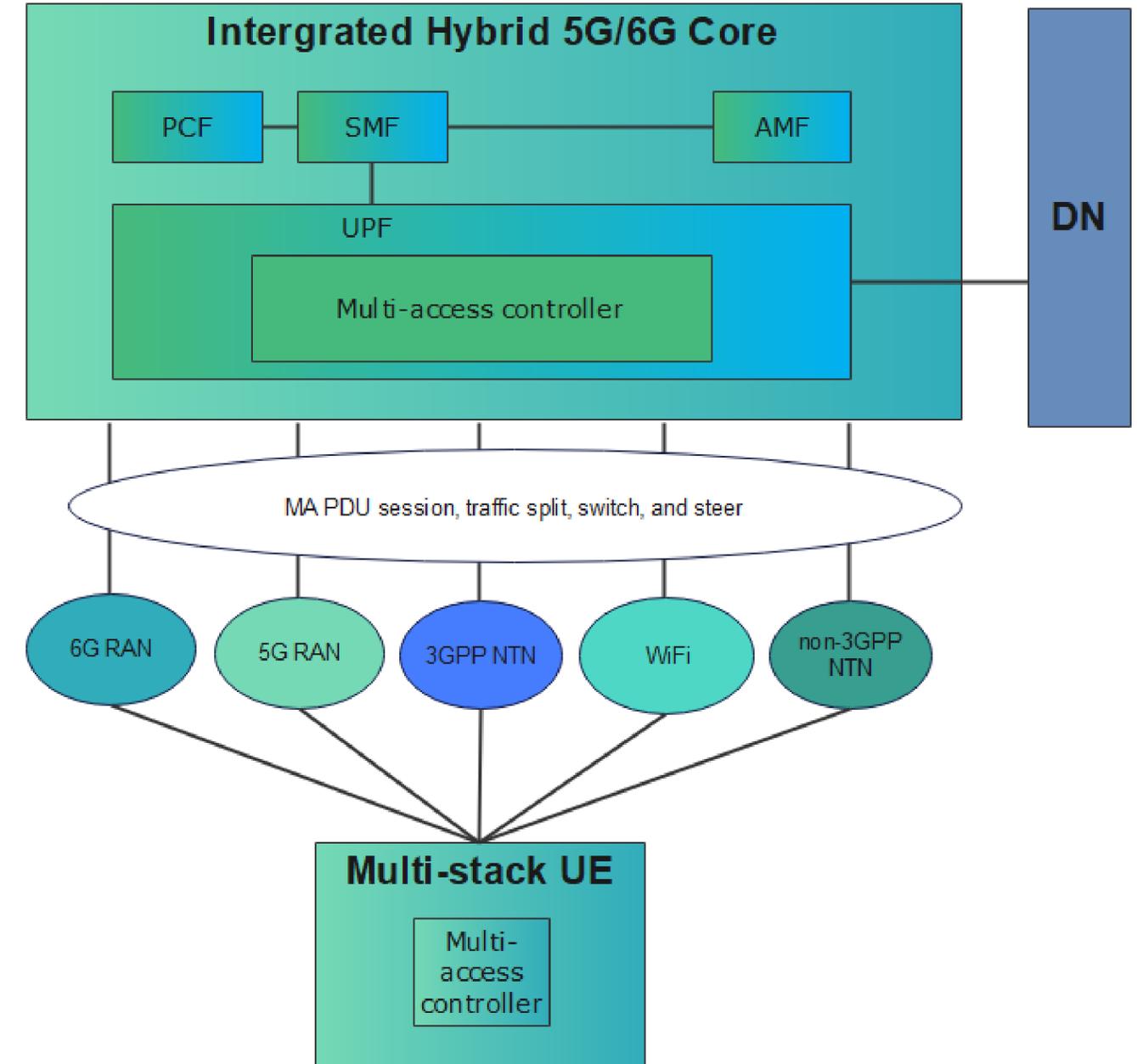
- Support external connectivity of subnetworks
- Security & Charging
- Bulk resources scheduling for subnetworks

Autonomous Sub-Networks

- Controlled by **Sub-Network Cluster Header** (Sub-NW CH), which is **either xNB or UE**, and could be deployed in **human body, smart home, UAV, vehicle, terrestrial xNB, or even satellite, etc.**
- Manage granted bulk resources within sub-networks
- Route local traffic within sub-networks to meet extreme performance.
- Route non-local traffic to macro network as relay
- Group management/mobility of UEs within sub-network
- Reduced device cost/hardware complexity by unified L1/2 design
- Can survive in case of loss of connectivity to Macro Network

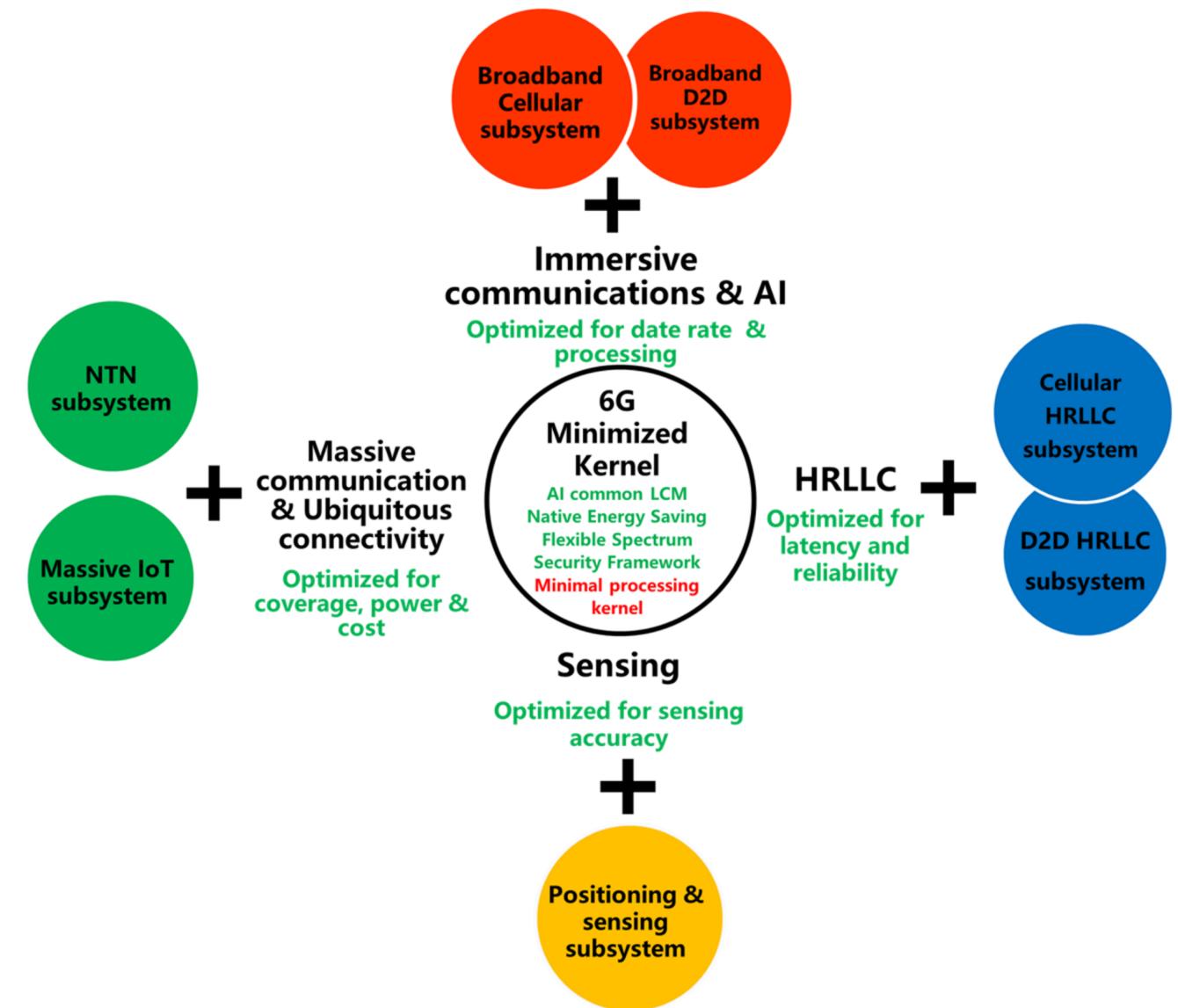
5G-6G Migration/Interworking

- Two options: hybrid 5G/6G core, or separate 5G/6G core to support 5G-6G migration
- The 5G/6G Hybrid Core: enabling Multi-Access flexibility accommodating various RAN
- The access technologies could be 5G and 6G, 3GPP or non-3GPP, and TN or NTN
- The UE, depending on its capabilities and form factor (cell phone, smart car, robot, etc.) may connect to one or more access technologies and/or core networks to achieve:
 - Different services
 - Extremely high data rate
 - Extended coverage
 - Effortless path switch
- All may be done beyond RAN and provide significant flexibility
- The baseline is a single-SIM UE



Network operation: Minimized Kernel + N subsystem

- 6G system should be like "one minimized kernel + N subsystem". Specifically the design of the versatile 6G system with minimized kernel mainly includes the following elements:
 - A minimized kernel provides common capabilities such as native AI, security, flexible spectrum management and minimal processing kernel
 - One or multiple subsystems are designed for each capability. Key technologies can be selected independently
 - Different subsystems can adopt different standard evolution cycles according to different market requirements. No inter-compatibility between subsystems are necessarily needed
 - The 6G system achieves flexible, low-cost, and low-power support for multiple subsystems by fast switching between the "Minimal processing kernel" in the minimized kernel and the "Complete processing core" of each subsystem



6G Minimized kernel + N sub-system

(<https://www.oppo.com/content/dam/oppo/common/mkt/footer/OPPO-6G-WhitePaper-EN.pdf>)

Proposal of study planning

- In order to make the Stage-1 study efficient, targeting to provide valuable output when Stage-2 SID starts, it is proposed to agree multiple SIDs or one SID with building blocks covering 6 directions:
 - eMBB,
 - Computation+AI (proposed objective in S1-241120),
 - Sensing,
 - Vertical,
 - Ubiquitous connectivity,
 - Network operation.
- When inputting use cases, each use case should indicate which direction(s) it addresses.
- An intermediate conclusion needs to be made for each direction before June 2025. It will be used for the guidance for stage-2 study.

Thank you

oppo

Example scenario-1: NET4AI (Mobile AI)

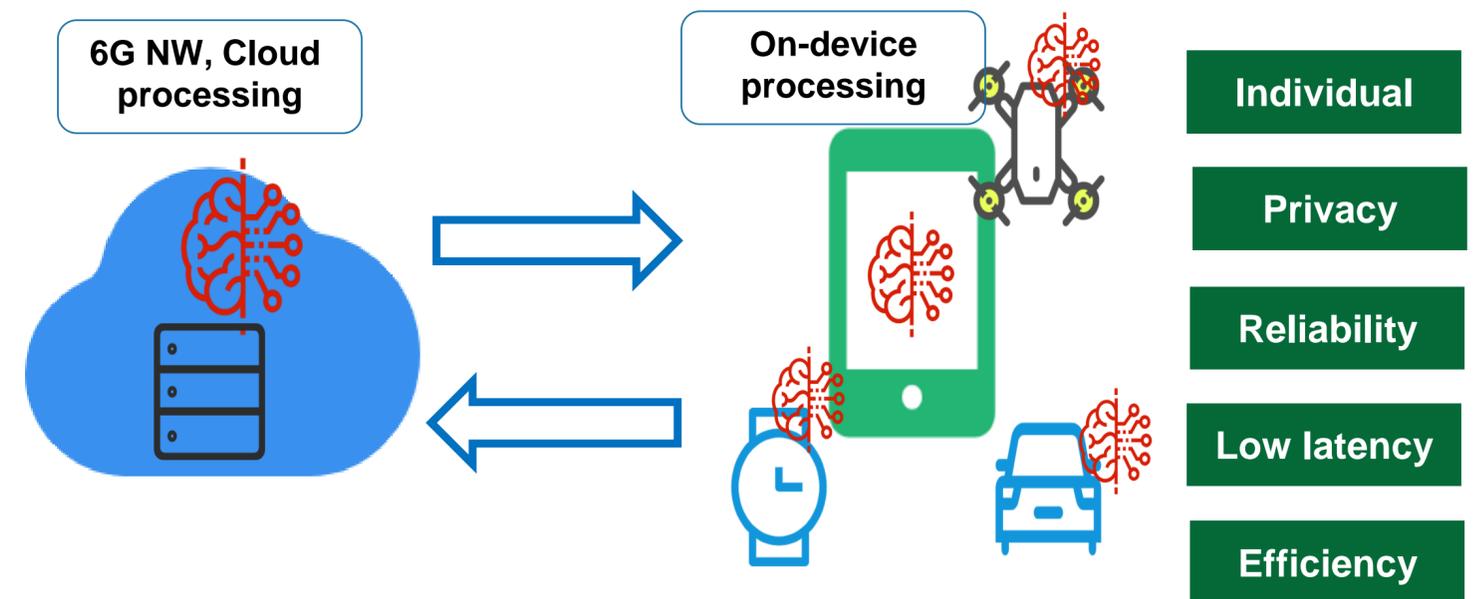
■ Mobile AI with device-NW/Cloud collaboration is the future of AI application

- The significant improvement in on-device computing power and efficiency is an important guarantee for the AGI large model. The near-future processor is expected to support local running of large-scale models with up to tens of billions of parameters;
- AGI model is "NW/cloud-side training and device-side deployment"

■ Mobile AI has the following advantages

- **Individual:** Create "user-level granularity" AI agents for users.
- **Efficiency:** Local reasoning can make full use of the contextual information of the terminal for reasoning
- **Privacy:** In many cases, data does not leave the local area for processing. In some scenarios, it needs to interact with the cloud for collaborative processing
- **Reliability:** inference using Personal device, user shares exclusive computing resources on device
- **Low latency:** Local processing does not depend on the stability of the transmission on air interface, and the inference delay is guaranteed

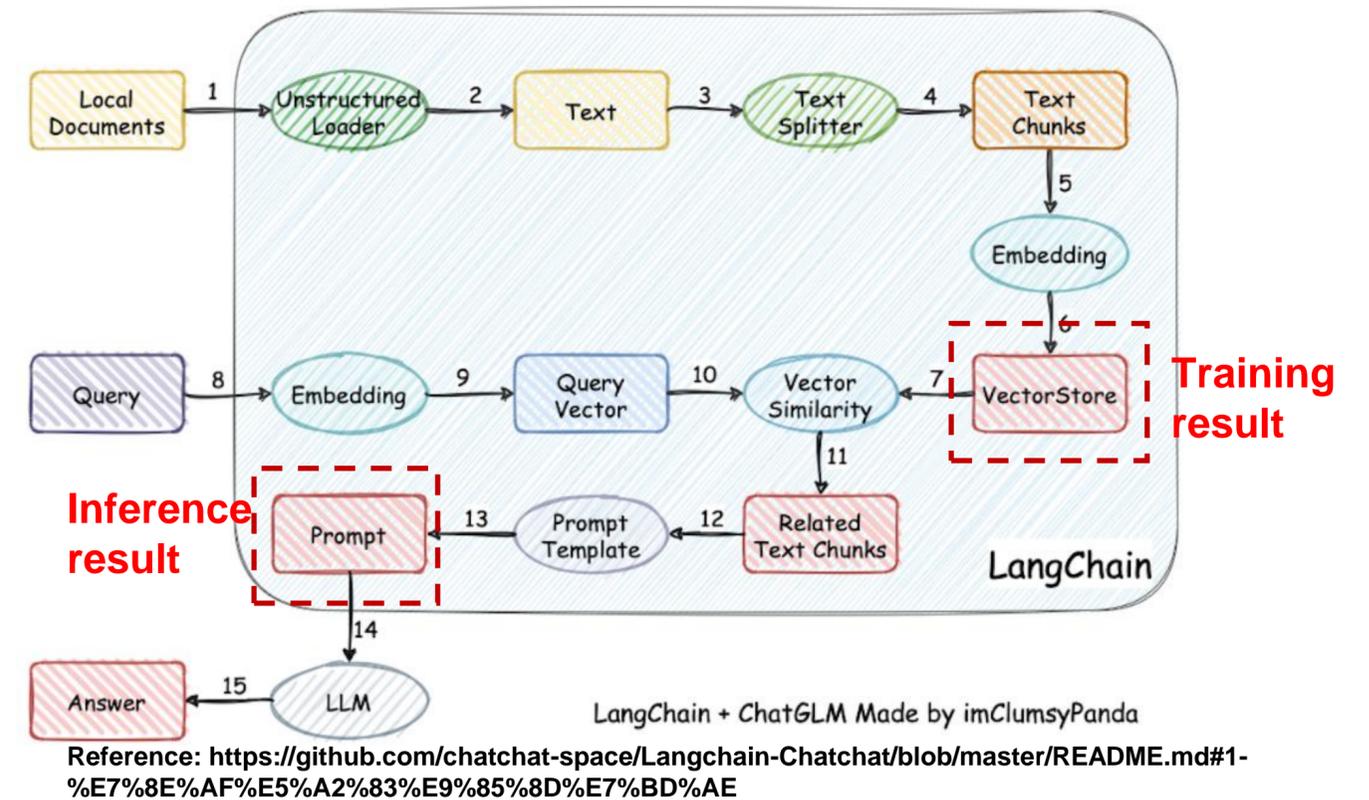
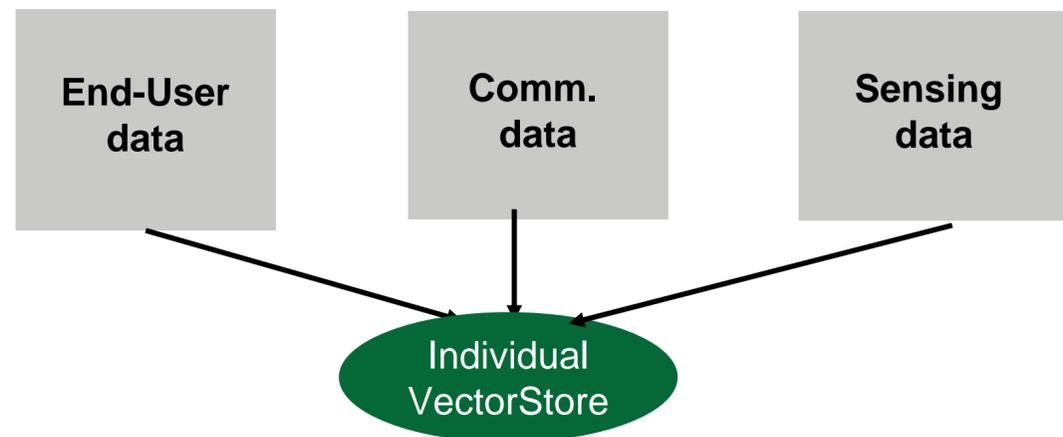
--



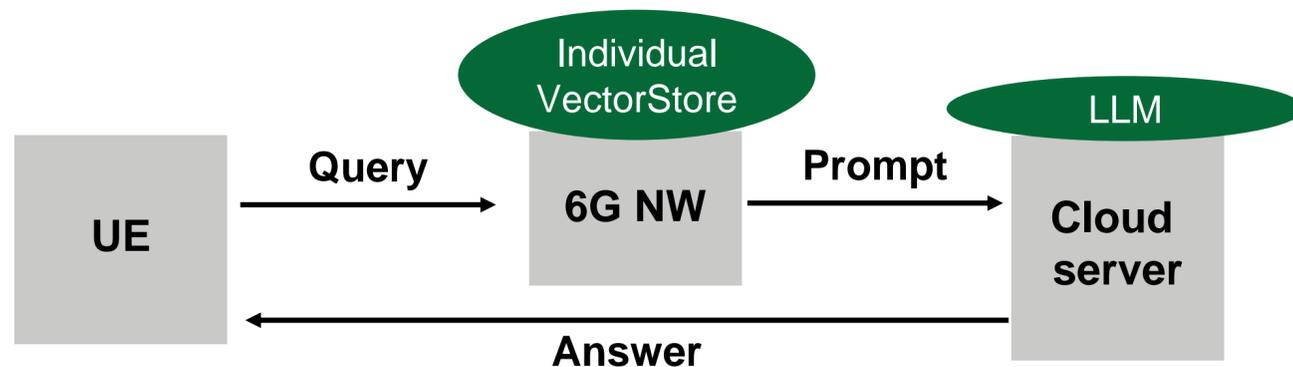
Mobile AI is to process data closest to its source, and supplement NW/cloud processing

Example UC-1: 6G NW Runs individual models for LLM

- **Training for RAG** (Retrieval-augmented generation (RAG) is a technique for enhancing the accuracy and reliability of generative AI models with facts fetched from external sources) Reference: <https://blogs.nvidia.com/blog/what-is-retrieval-augmented-generation/>



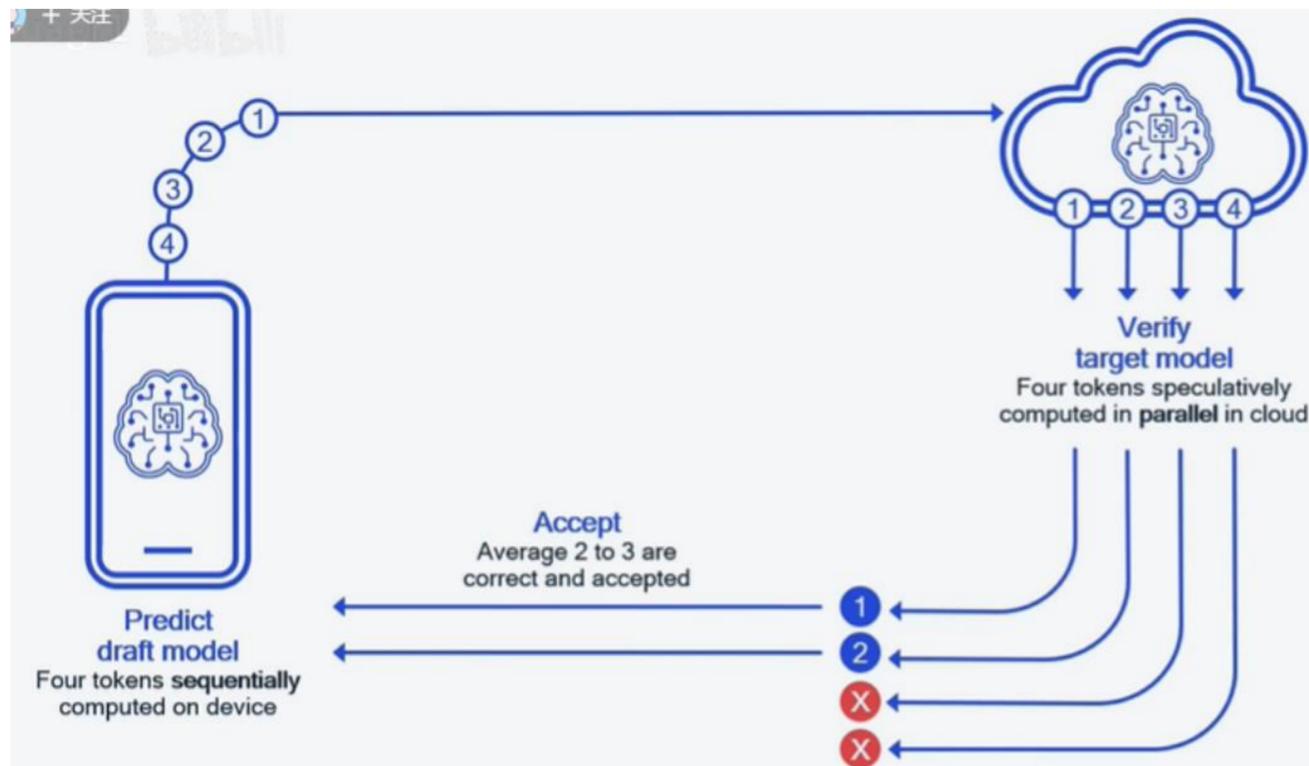
- **Inference for individual language chain (step 7-15 in LangChain)**



■ 3GPP network shall be able to support:

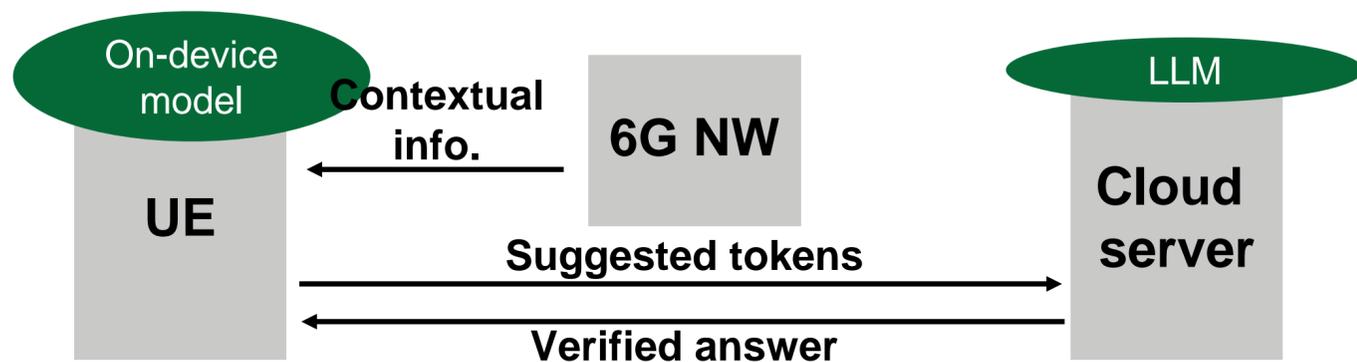
- **Data**
provide sensing/comm. data for model training and/or inference
- **Computation**
support the inference for generating prompts based on vector store
The e2e KPI should include communication + computation
- **Connection**
support “On-path” content awareness and processing
- **Model**
The individual vector store is maintained in 6G network per UE

Example UC-2: Hybrid AI inference



- The UE assists in inference, and the cloud confirms or modifies the inference results and returns them to the terminal. The round-trip delay based on the amount of information determines the delay required for application-side computing power.

□ Hybrid UE-Cloud inference



- **3GPP network shall be able to support:**
 - **Data**
Contextual information needs to be exposed to UE
 - **Connection**
Efficient/direct information exposure to UE
Round trip KPI for suggested-verified token

Example UC-3: Cross-domain AI collaboration

- During roaming across geographical regions, UE may traverse different RANs and CNs, or switch to App Servers located in other regions. This dynamic environment poses challenges to maintaining consistent user quality of experience.
- Cross-domain AI collaboration can integrate information and AI capabilities from various domains (such as CN domain, RAN domain, Application domain, Management domain, etc.), enabling a seamless and consistent high-quality experience for roaming users.

