

## 5.5G Empowered Mobile AI Economy

**Liu Haoyi (Antonio)**  
**Solution Architect**  
**Huawei Technologies**  
**Nov. 6<sup>th</sup>, 2024**



# AI Phone Application at a Glance

Circle to search  
(in video/graphic)

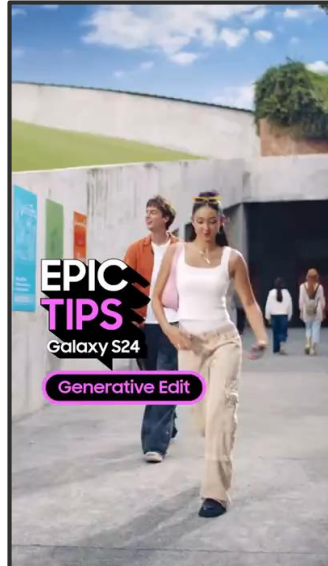


Webpage Summary  
(from webpage)

|  |  |
|--|--|
| Official name                              | Statue of Liberty National Monument, Ellis Island and Liberty Island |
| Designated                                 | May 27, 1971   |
| Reference no.                              | 1535 <sup>[5]</sup>  |
| New York State Register of Historic Places |  |
| Designated                                 | June 23, 1980 <sup>[6]</sup>   |
| Reference no.                              | 06101.003324   |
| New York City Landmark                     |  |
| Type                                       | Individual   |
| Designated                                 | September 14, 1976 <sup>[7]</sup>                                    |
| Reference no.                              | 0931   |

The statue is a figure of Libertas, the Roman goddess of liberty. She holds a torch above her head with her right hand, and in her left hand carries a *tabula ansata* inscribed JULY IV MDCCLXXVI (July 4, 1776, in Roman numerals), the date of the U.S. Declaration of Independence. A broken chain and shackle at her feet as she walks forward, commemorating the national abolition of slavery following the American Civil War.<sup>[8]</sup> After its dedication the statue became an icon of freedom and of t<sup>h</sup>e United States, being

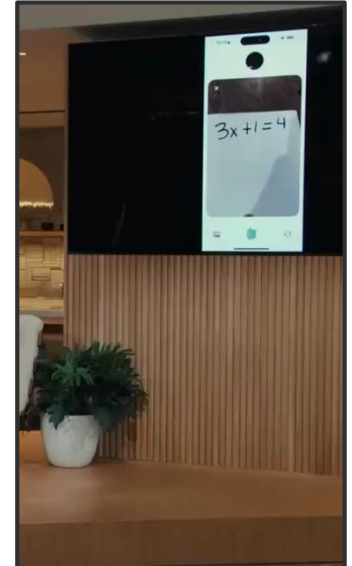
Generative Edit  
(in video/picture)



iPhone16 Siri  
(Multi tasks)



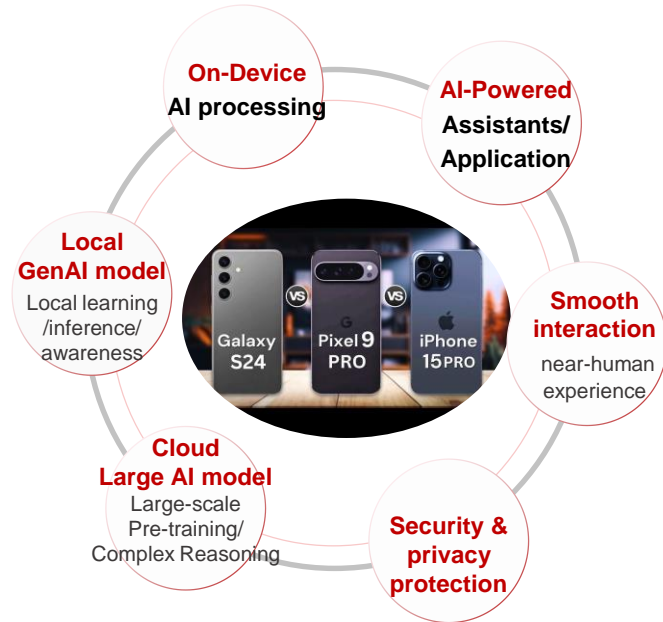
ChatGPT 4o  
(visual reasoning)



# GenAI Phone: GPU+ Native Support LLM for Both Local and Cloud

## What is “GenAI Phone” ?

A Smartphone Integrate Generative AI capability



View point collect from:  
2024 IDC, Counterpoint, OpenAI, Gartner...

## Key Feature: Added Support for AI Models

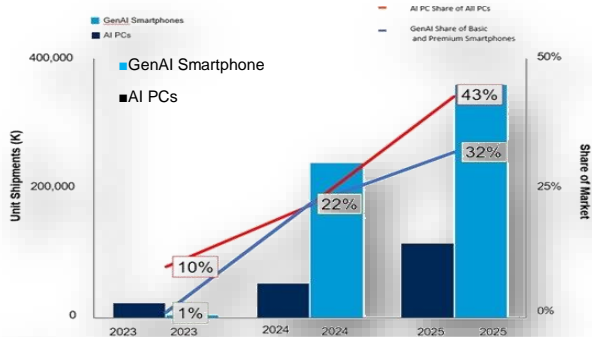
| Brand                         | Samsung S24 Ultra                      | iPhone 16 Series   | Google Pixel 9 Pro |
|-------------------------------|--|--------------------|--------------------|
| OS                            | Android 14                             | iOS 18             | Android 14         |
| chipset                       | Snapdraon 8 Gen3                       | A18 Bionic         | Tensor G4          |
| NPU (TOPS)                    | 48                                     | 35                 | 38                 |
| AI/LLM Model                  | Samsung AI                             | Apple Intelligence | Gemini             |
|                               | <b>On device local + Cloud Synergy</b> |                    |                    |
| On device model size(Billion) | N/A                                    | 3B*                | 2.2B*              |
| DRAM (GB)                     | 12                                     | 8                  | 16                 |

\*B: Billion parameters

# GenAI Phones are Rapidly Growing

## High Shipment Volume

**240 million shipped**  
by the end of 2024

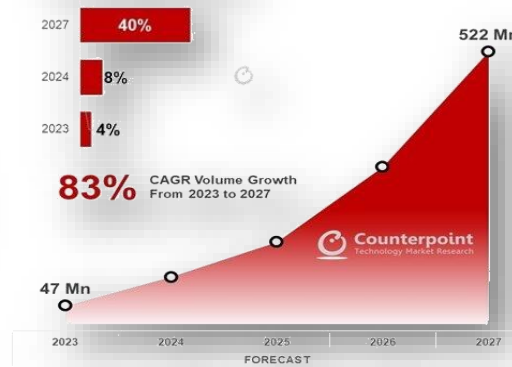


Gartner

## High Shipment Share

**8% @2024, 40% @2027**

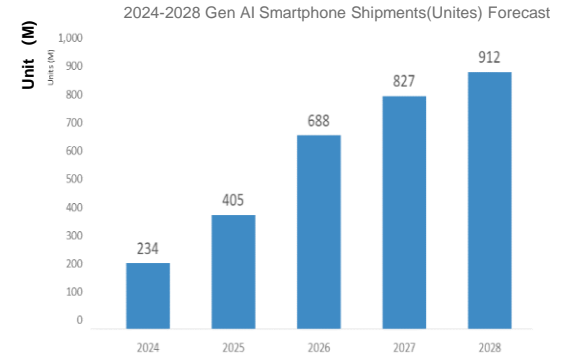
Share of GenAI Smartphones,  
Global Smartphone Market



Counterpoint

## Rapid Shipment Growth

**78.4% CAGR**  
@ 2023~2028



IDC

# GenAI UCs Continues to Develop Rapidly **With More UCs Relying on Cloud**

## Local Computing

**Simple Word** Processing with **Low latency**



**Web translation**



**Voice call interpreter**



**Voice-to-text conversion**



## Cloud Computing

**Complex Graphics** Processing with **High Throughput**



**Eliminate /Expand**



**Circle to search**



**Summary**

| UC           | AI Function             | Latency          | Throughput |
|--------------|-------------------------|------------------|------------|
| Summary      | Voice-to-text           | <b>Real-time</b> | 1~100kbps  |
| Web Browsing | Web page translation    | <b>Real-time</b> | 1~100kbps  |
| Calling      | Interpretation          | <b>Real-time</b> | 100k~1Mbps |
| Notebooks    | Handwriting recognition | <b>Real-time</b> | 1~100kbps  |
|              | Text translation        | <b>Real-time</b> | 1~100kbps  |
| Social       | Reply suggestions       | <b>Real-time</b> | 1~100kbps  |

| UC           | AI Function           | Latency | Throughput         |
|--------------|-----------------------|---------|--------------------|
| Photo edit   | Character elimination | second  | <b>1~10Mbps</b>    |
|              | Picture edit          | second  | <b>1~10Mbps</b>    |
| Summary      | Recording summary     | second  | <b>1~10Mbps</b>    |
| Web Browsing | Web page summary      | second  | <b>1~10Mbps</b>    |
|              | Circle and search     | second  | <b>1~10Mbps</b>    |
| Calling      | Record to summary     | second  | <b>1~10Mbps</b>    |
|              | Video Optimize        | minutes | <b>&gt;100Mbps</b> |



**"Smart assistant" & "office creation" GenAI UC**

**Can NOT** be implemented only by the local computing, need to be implemented by **device-cloud collaboration**

# Device-cloud Synergy is the Mainstream Solution of GenAI Phones

Major AI function are computed on Cloud

| UC           | AI Function             | Works on |
|--------------|-------------------------|----------|
| Photo        | Character elimination   | Cloud    |
|              | Picture Expansion       | Cloud    |
|              | Picture edit            | Cloud    |
| Recording    | Voice-to-text           | Device   |
|              | Recording Summary       | Cloud    |
| Web Browsing | Web page Summary        | Cloud    |
|              | Web page translation    | Device   |
| Calling      | Circle and Search       | Cloud    |
|              | interpretation          | Device   |
| Notebooks    | Record to Summary       | Cloud    |
|              | handwriting recognition | Device   |
|              | Text translation        | Device   |
| Social       | Note Summary            | Cloud    |
|              | Reply suggestions       | Device   |

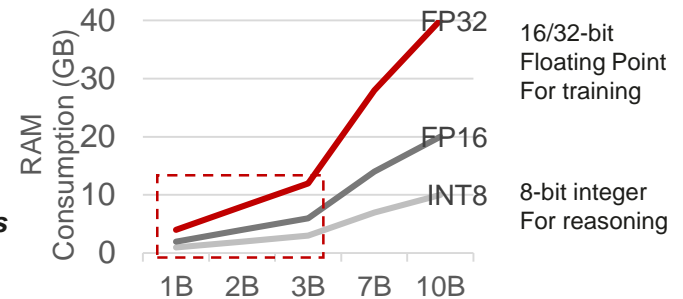
On-Device AI Model  
**Consuming Energy**

*1 picture consumes 5% phone battery*

On-Device AI Model  
**Consuming Memory**

*3 Bn parameters consumes 3~10GB RAM*

| Task                   |   | kwh    |
|------------------------|---|--------|
| Text classification    | ✓ | 0.0002 |
| Graphic classification | ✓ | 0.0007 |
| Object identification  | ✓ | 0.0038 |
| Text generation        | ✓ | 0.0047 |
| Abstract               | ✓ | 0.0049 |
| Graphic generation     | ✗ | 0.2907 |



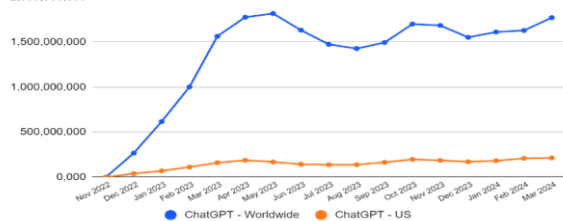
# Generative AI with multi-modal, **Mobile UL Traffic Expected to Increase 7~9X**

## OpenAI: 13% Month-on-Month Traffic Growth in the past 24 month

ChatGPT - Worldwide and US Traffic

Desktop & Mobile Web Visits

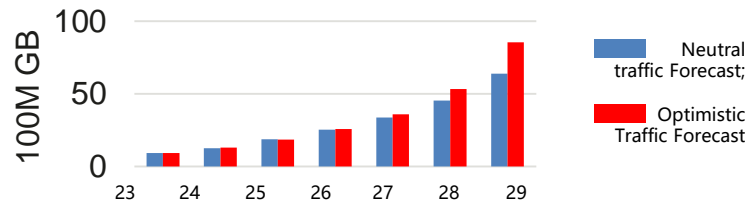
2,000,000,000



Source: OpenAI

7~9x UL traffic will be created via Mobile AI from chat pat, to pilot, to agent

UL traffic prediction 2025-2029



Source: Huawei xLab

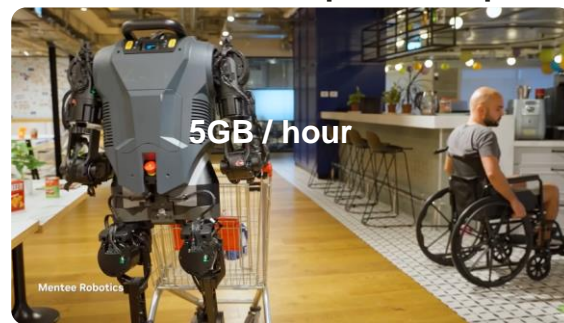
voice dialog



+ Visual intelligence



+multi-lens camera + Spatial computation



# AI Full-modal Interaction Revolution **Raises Higher Demands on Network**

UL  
Large  
Throughput

**20-30  
Mbps**

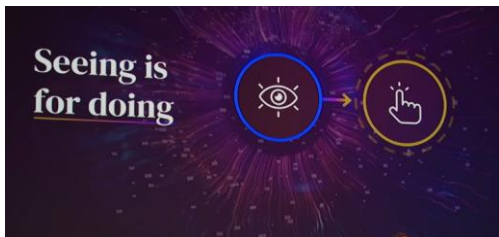
Low  
Latency

**20-25 ms**

UL  
Large  
Capacity

**7 - 9 X**

Improved  
Interaction Quality



Audio and video are fully modal  
(space video..)

AI 3D 8K video (2 minutes): 1.1 GB  
2D 480p video (2 minutes): 12 MB

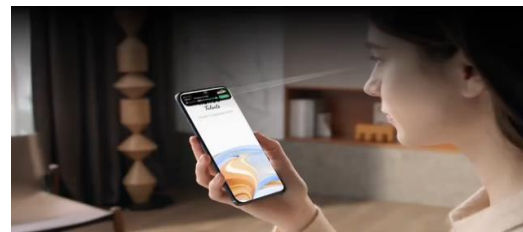
Improved  
Interaction Efficiency



Content retrieval → Content generation

Document retrieval time: minutes  
AI searches and generates content: seconds.

Improved  
Interaction Objects



people-models Interaction  
& inter-model interaction

Upload training row data: TB  
Download / update model data: GB



# AI + IoV: Reshaping driving and the value of connectivity

New wireless-enabled vehicles: >80 mn/year

Intelligent connected vehicles: Safer, easier, and more enjoyable



Intelligent cockpit penetration: >90% by 2030



5-10 HD screens /vehicle >Gbps DL experience

Experience value

10x video streamed; 100x data consumed by vehicles by 2026

Intelligent driving penetration: >40% by 2030



100GB /vehicle/month <20ms E2E latency

Time and security value

Traffic speed: ↑20%; accidents: ↓25%

New value of connectivity

萝卜  
快跑

萝卜快跑自动驾驶 bilibili

Apollo Technology Day

# China L4 Robotaxi: Remote Assistance Uplifting UL Demand

**Remote Takeover @ Anytime, is the prerequisite for secure commercial use**



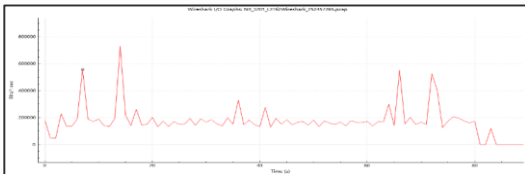
## Regulation:

- **3** vehicles need **1** remote safety officer
- **Real-time video surveillance** & record
- When **accident or fails**, previous **90s video data** uploaded

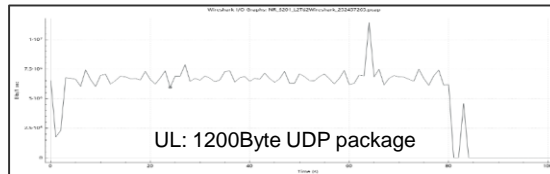
## Data transmission requirements:

- **AI driving**: front camera real time monitoring
- **Remote takeover**: **1**(in-car) + **5**(on-car) **cameras**, concurrent working and transmitting
- **Accident record**: full buffer upload all data, 90s records= 2 GB

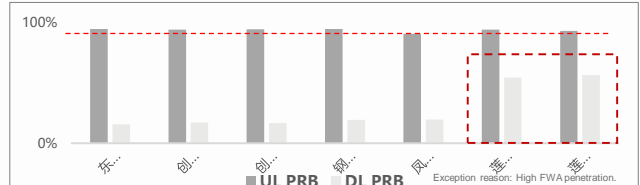
**Downlink: avg. 180Kbps, peak 720Kbps**



**Uplink: avg. 6.5Mbps, peak 11.5Mbps**

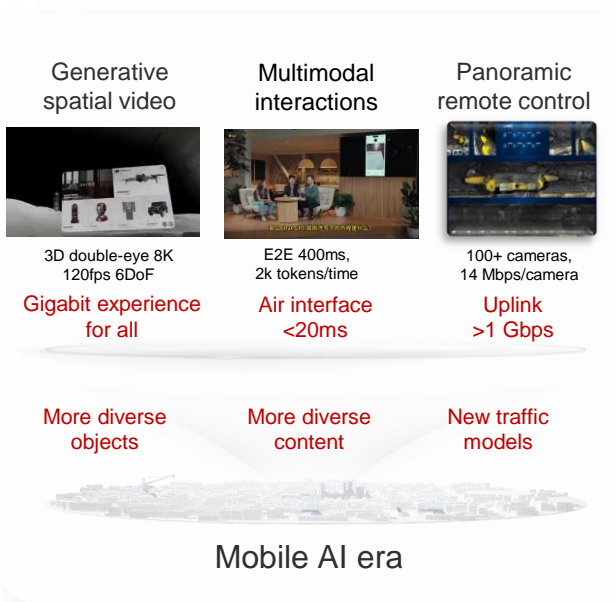


**Log File Upload During Charging, UL PRB > DL**



# Networks for AI, Driving faster 5G-A network deployment

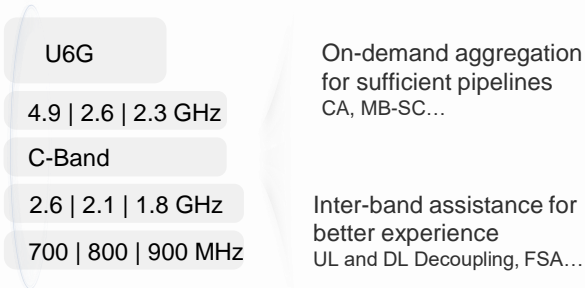
## Mobile AI: Higher network requirements



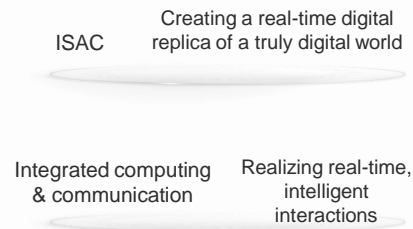
## Faster 5G-A deployment laying a foundation for mobile AI

IoP and IoV: All frequency bands support 5G-A for ultimate experience

Pooling frequency bands

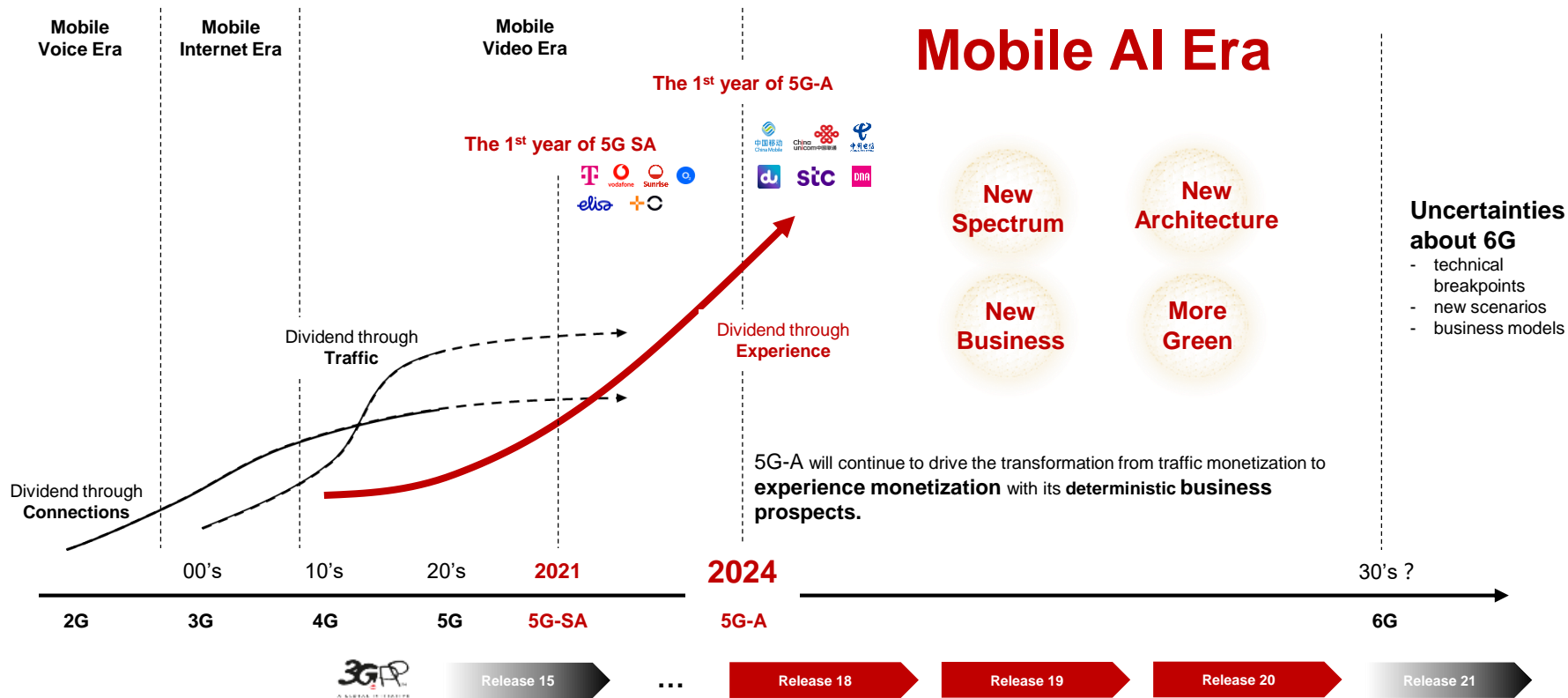


IoE-ready 5G-A with integrated sensing, computing, and communication



Gigabit experience for all    10 Gbps for hotspots    Gigabit uplink    Ultra-low latency    ...

# Mobile AI Era is Coming, 5G-A Continues to Evolve, while 6G Sets Higher Goals





Thank you.

**Bring digital to every person, home and organization for a fully connected, intelligent world.**

