December 2022 @QCOMResearch

Transforming the Connected Intelligent Edge to Realize 5G's Full Potential

Leading wireless innovation for more than 35 years

Digitized mobile communications

Analog to digital

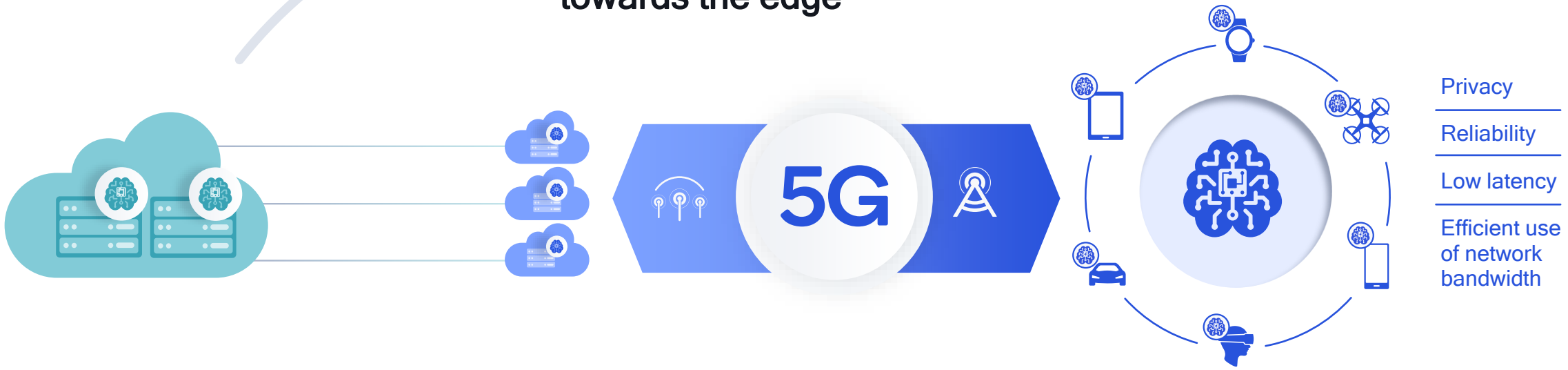
Redefined computing

Desktop to smartphones

Transforming industries

Connecting virtually everything

To efficiently scale, AI processing is expanding towards the edge



Qualcomm is leading the realization of the Connected Intelligent Edge

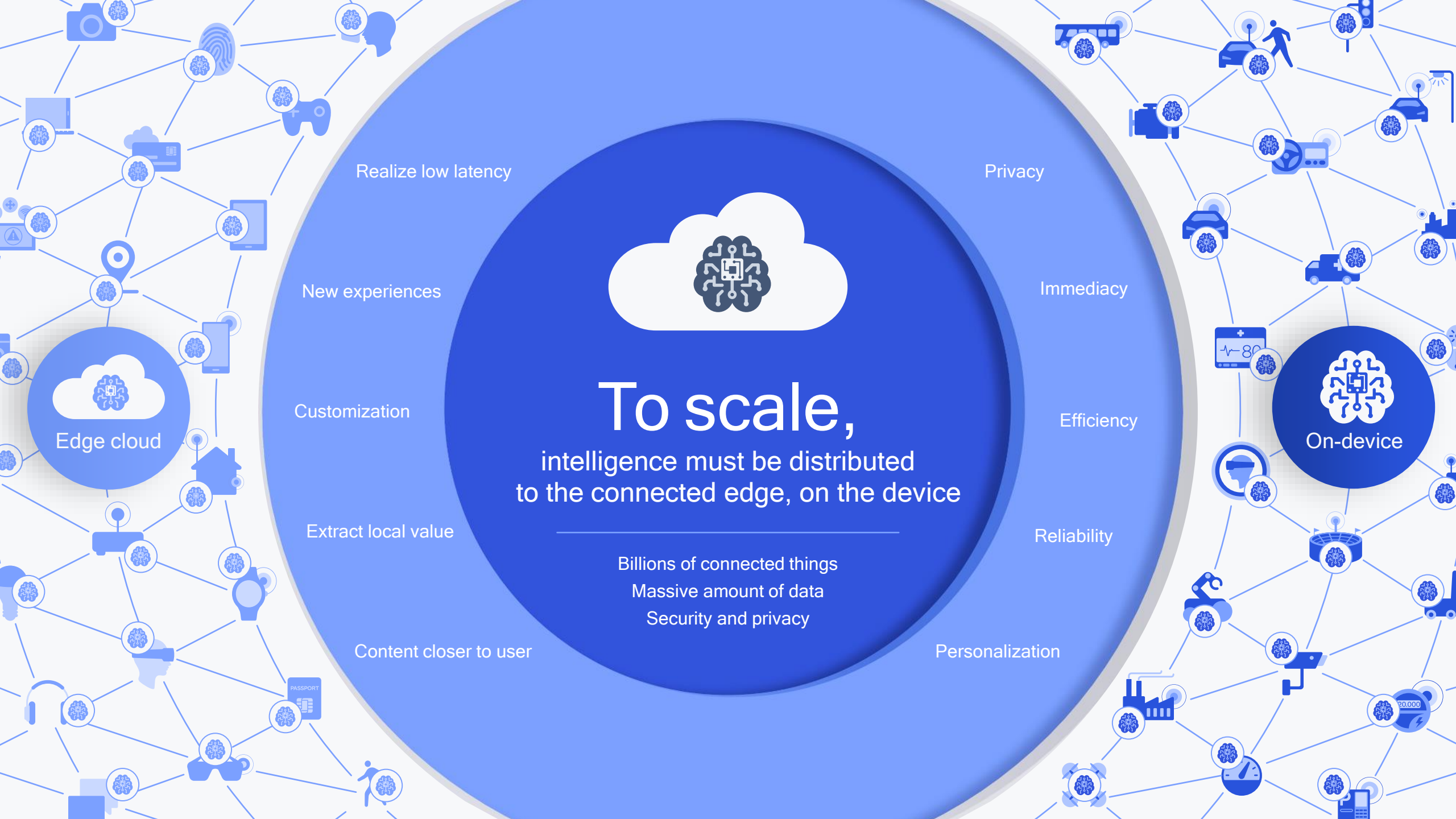
Convergence of:

Wireless connectivity
Efficient computing
Distributed AI

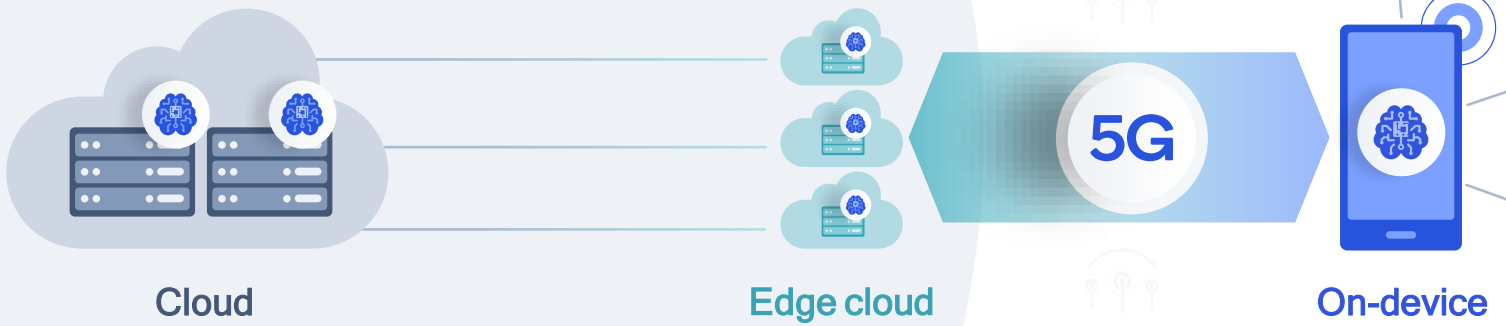
Unleashing massive amount of data to fuel our digital future

Today,

intelligence is primarily associated with the cloud


Realize low latency
New experiences
Customization
Extract local value
Content closer to user
Privacy
Immediacy
Efficiency
Reliability
Personalization
Edge cloud
On-device
To scale,
intelligence must be distributed
to the connected edge, on the device
Billions of connected things
Massive amount of data
Security and privacy

Efficiently scale intelligence in the era of the connected edge



Processing closer to devices at the edge derives new system values (e.g., lower latency, enhanced privacy)

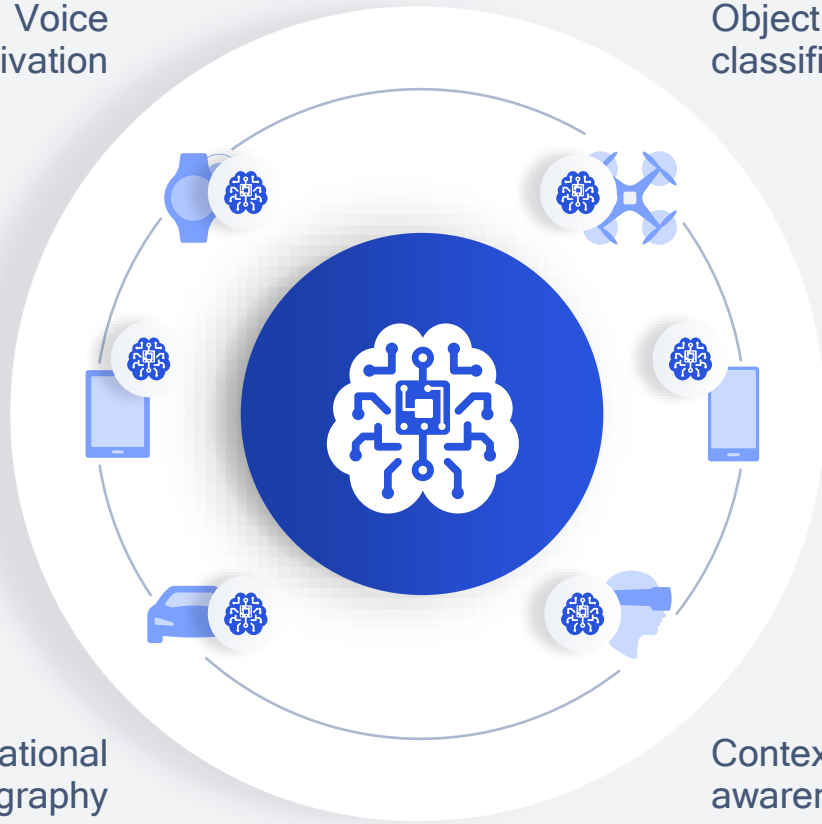
Past	Today	Future
Cloud-centric AI	Partially-distributed AI	Fully-distributed AI
AI training & inference in the central cloud	Power-efficient on-device AI inference	With lifelong on-device learning

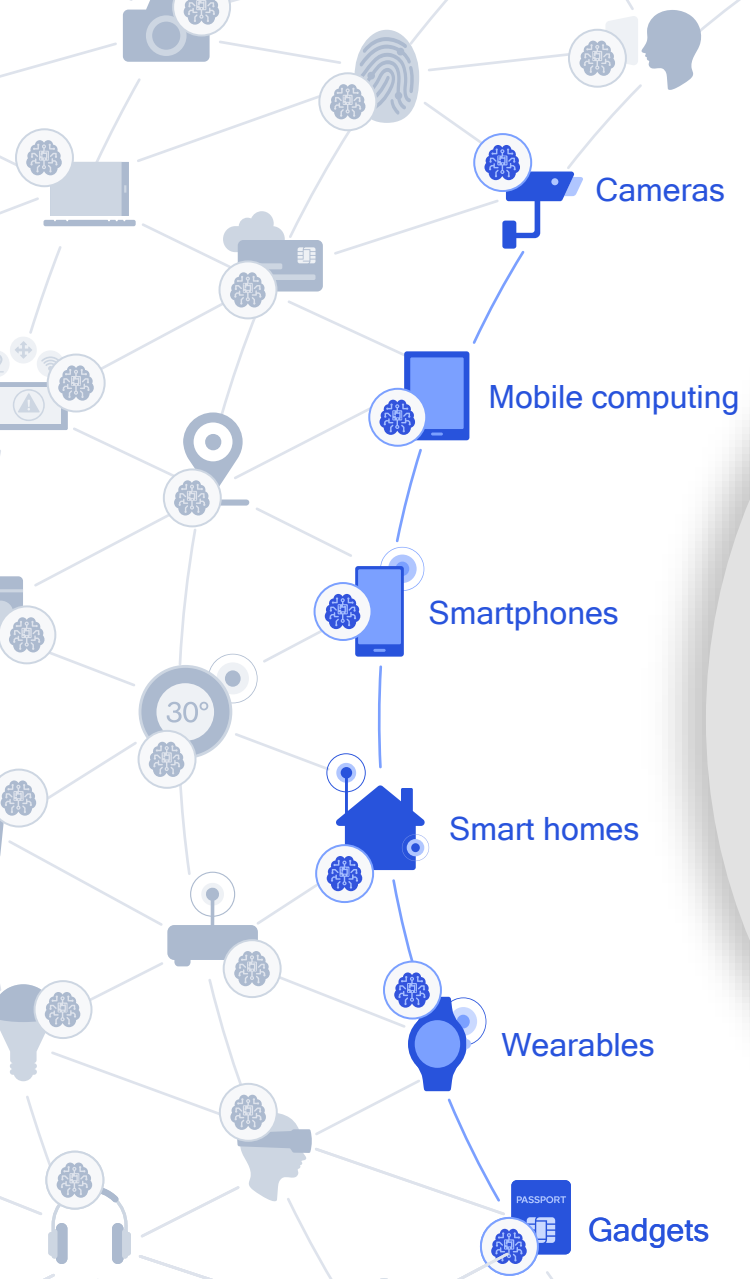
- Local network analytics
- Low-latency interactive content
- Boundless XR
- On-demand computing
- Industrial automation and control
- Enterprise data

Connected Intelligent Edge

brings new and enhanced services

Edge Cloud + On-device AI





Mobile is becoming the pervasive AI platform bringing broader

Societal benefits

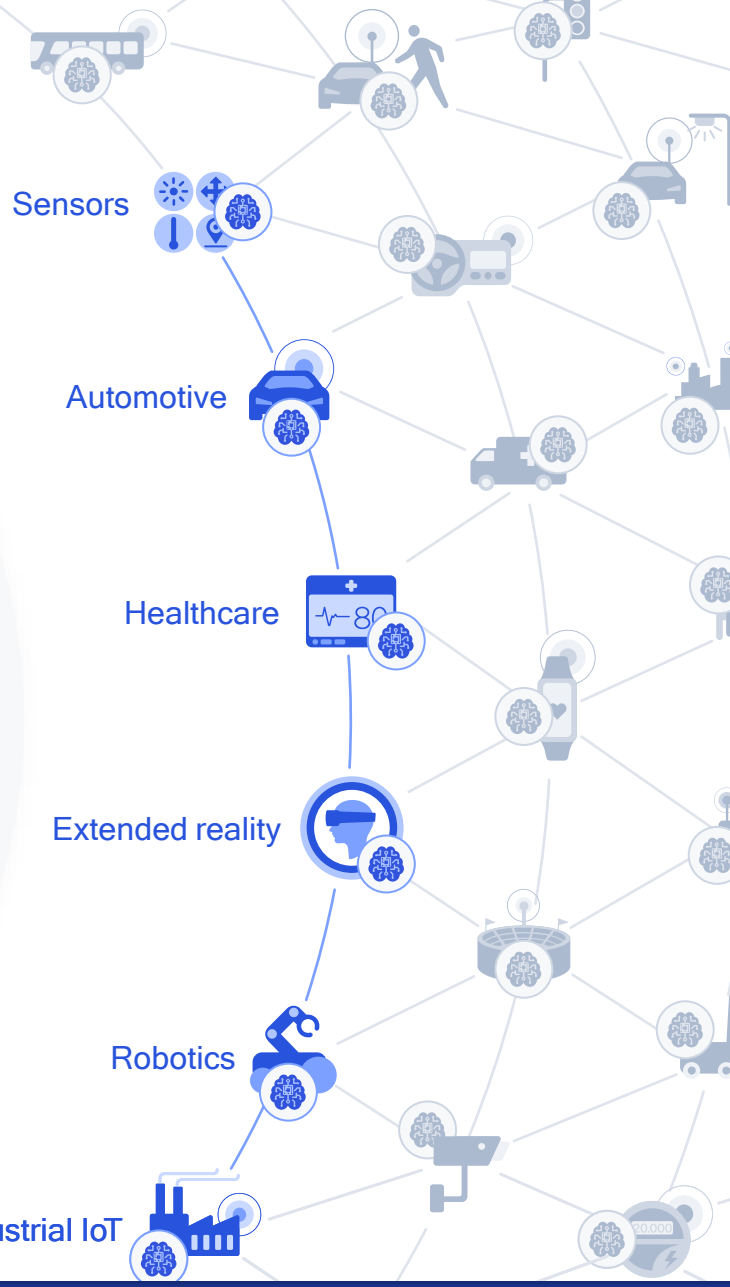
and the promise of ubiquitous on-device AI

Rapid replacement cycles

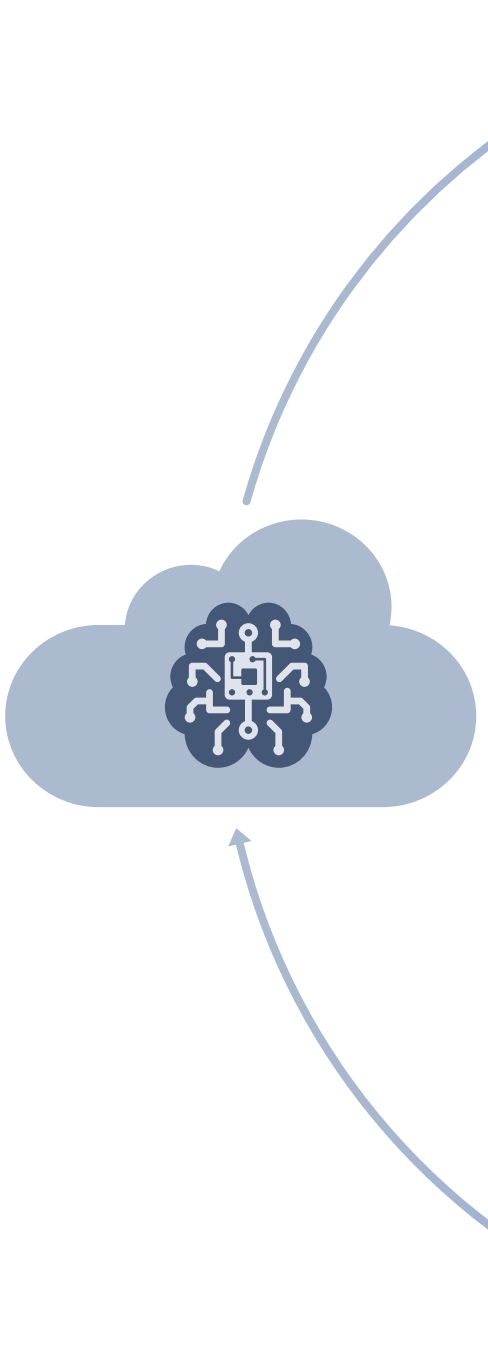
Superior scale

Power/thermal efficiency

Integrated/ optimized technologies

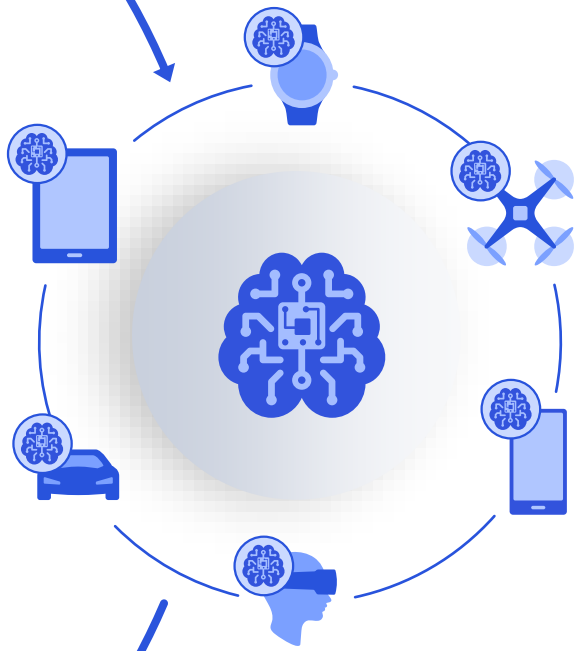


Driver safety	Virtual assistant	Enhanced security	Superior photography	Healthcare and more...



On-device intelligence is paramount

Process data closest to the source, complement the cloud



- Privacy
- Reliability
- Low latency
- Efficient use of network bandwidth

The challenge of AI workloads

- Compute intensive
- Large, complicated models
- Complex concurrencies
- Real-time
- Always-on

Constrained mobile environment

- Thermally efficient for sleek designs
- Requires long battery life for all-day use
- Storage/memory bandwidth limitations

Power and thermal efficiency

Critical to the promise of AI on a wide range of connected devices

Our technologies are transforming the connected intelligent edge today

Inventing technology at scale to realize the promise of massive on-device AI

Providing the connectivity fabric with LTE, Wi-Fi, Bluetooth today

Security

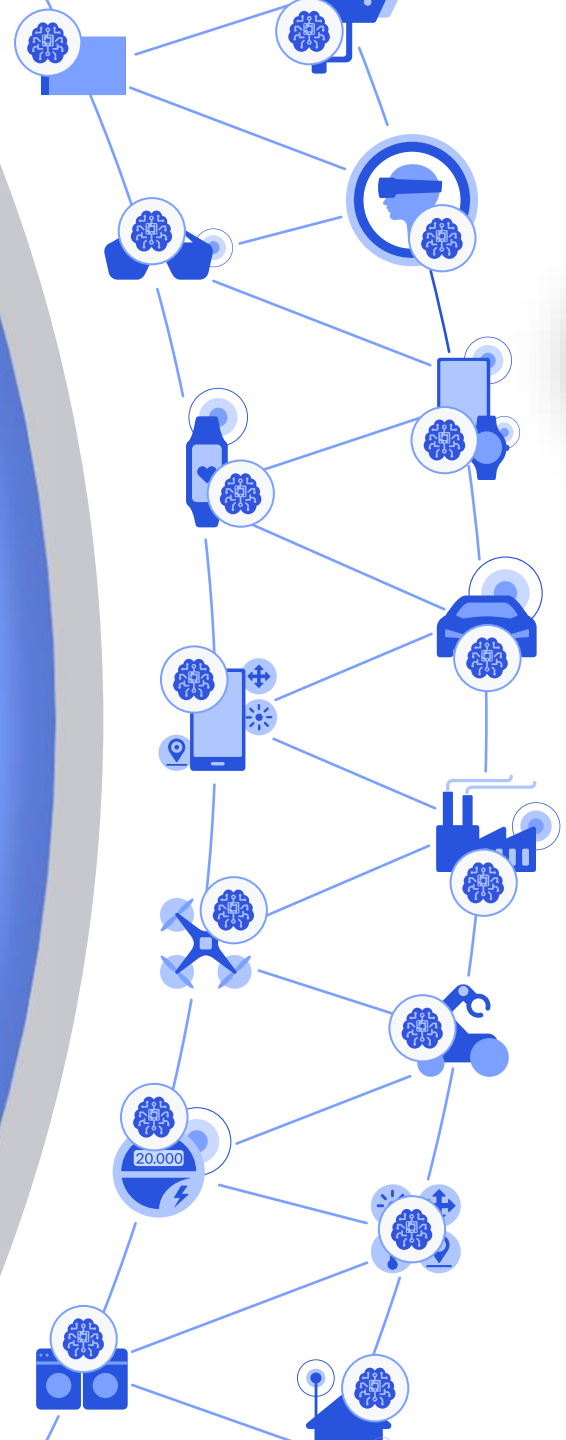
Sensing

CPU
DSP
GPU
Processing

Connectivity

Edge services

On-device AI



Perceive
Hear, see, monitor, and observe

Reason
Learn, infer context, and anticipate

Act
Act intuitively, interact naturally, and protect privacy

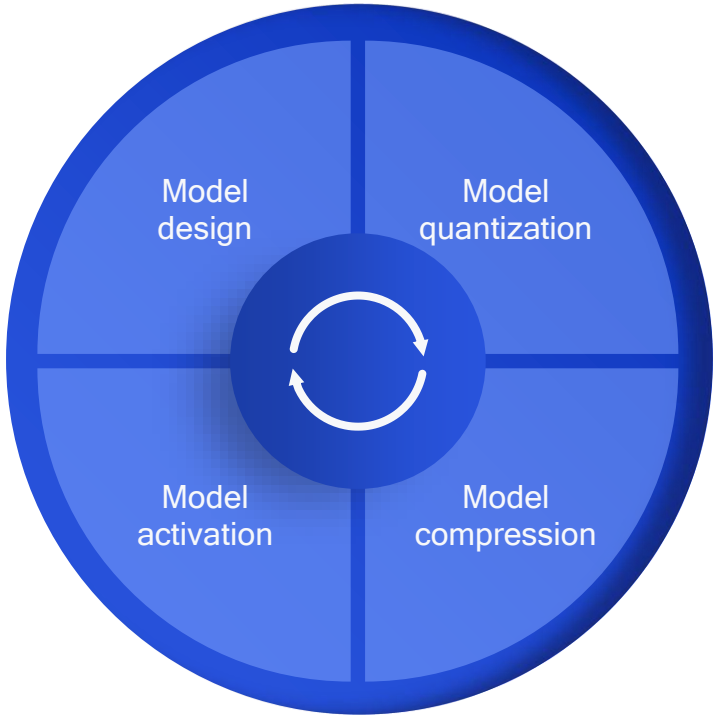
Qualcomm Technologies is making on-device AI ubiquitous

Efficient hardware

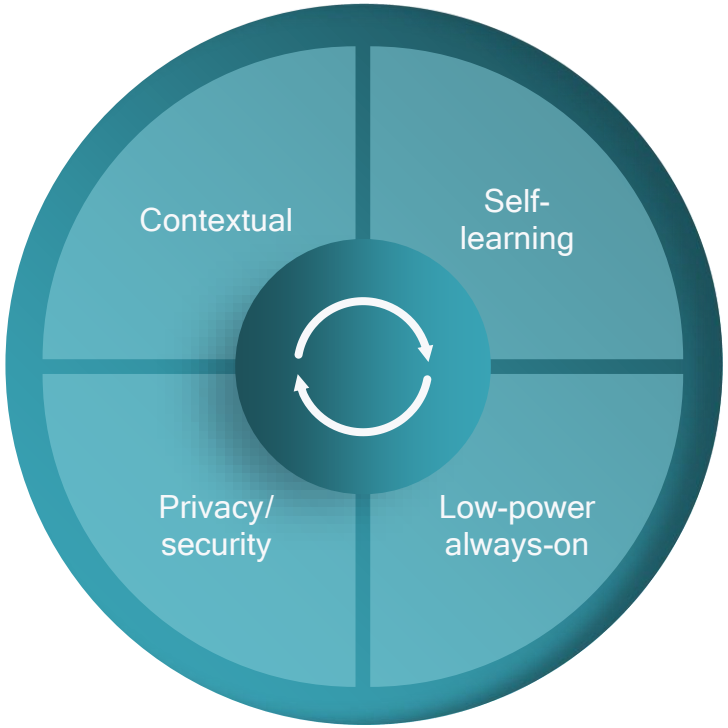
Algorithmic advancements

Software tools

Power efficient AI



Personalized AI



Advancing 5G to fulfill its full promise

An innovation platform for this decade and beyond

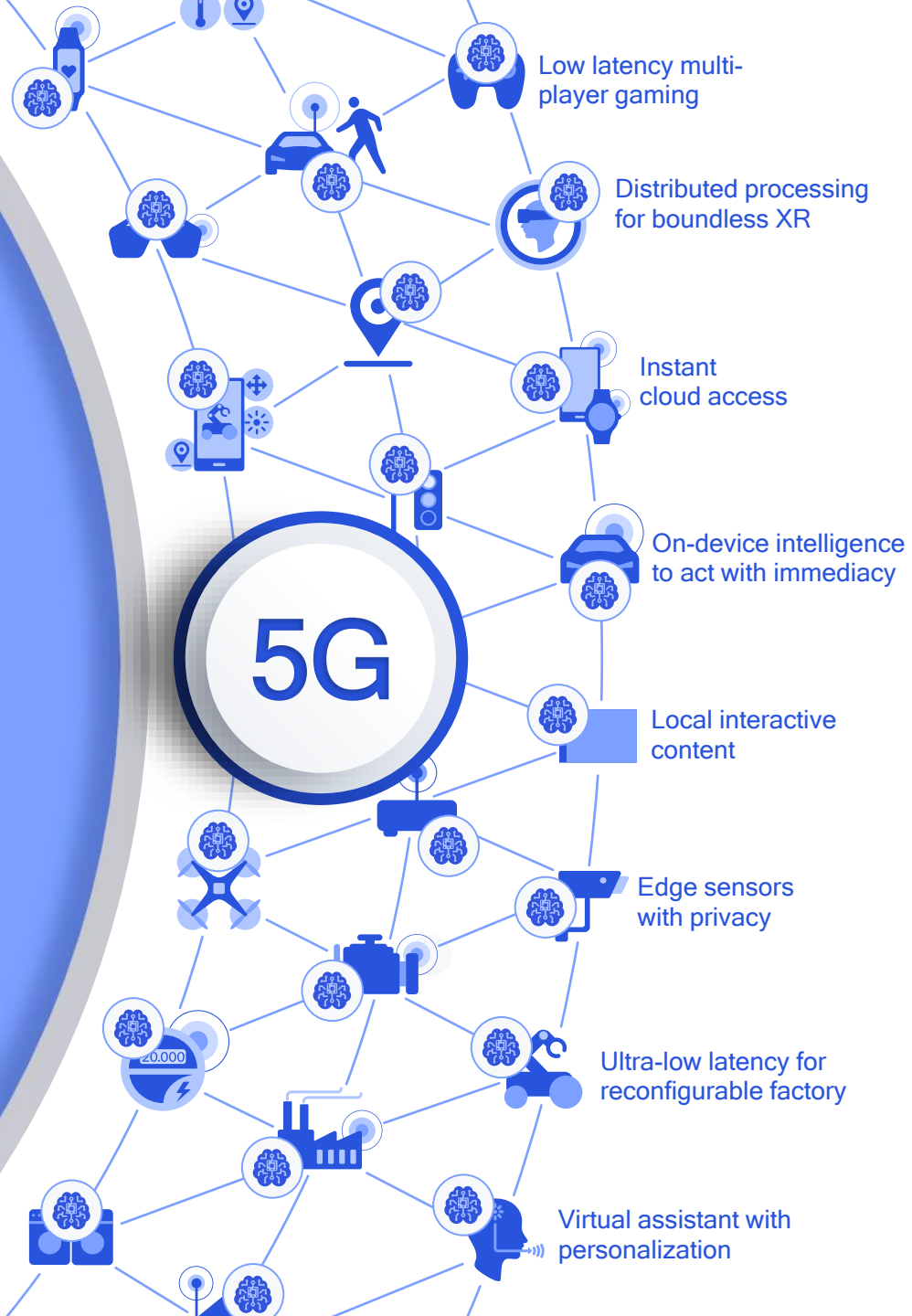
A unifying connectivity fabric for the connected intelligent edge

Connecting virtually everything, an innovation platform for future services

- Extreme mobile broadband
- Mission-critical services
- Massive Internet of Things

The connected intelligent edge transformation realizes the full potential of 5G

- New experiences with new levels of immersion, immediacy, personalization and privacy
- Creating new industries and transforming existing industries in the new era of distributed autonomy
- Essential on-device capabilities augmented with processing/compute, content, control,… at edge cloud



Enhancing services, creating new services, new deployment models with our end-to-end expertise

Cloud

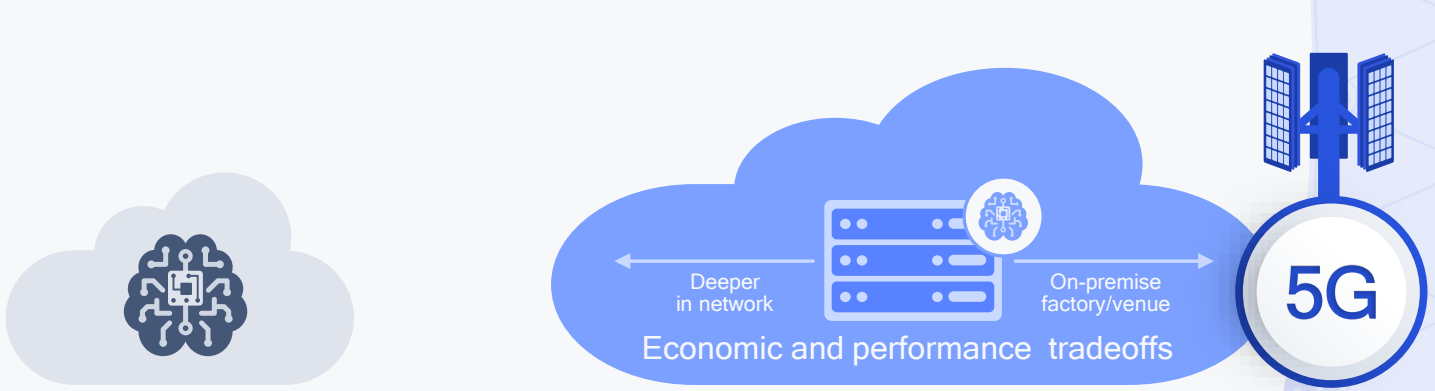
Big data, AI training, less delay sensitive content, storage,…

Augmented by edge cloud

Compute/processing, content, control, storage,.. closer to user[1]

Driving the best possible on-device capabilities

Sensing, processing, security, intelligence



Latency could be over 100s ms today

Latency as low as 1 ms

Cooperation between operators, infra vendors, OEMs, content providers, cloud providers,…

- Realize 5G's low latency
- Distributed/split functions over 5G
- Extract value – vs. 'dumb' data pipe
- Customization, local value
- New deployments, like private networks

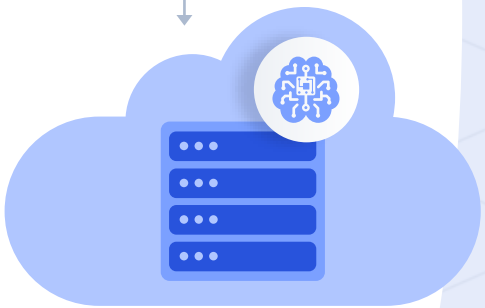
1. Such as distributed/virtualized core, distributed packet gateway functionality for low latency, mobile edge compute, related to MEC Multi Access Edge Computing as defined by ETSI

Enriched user experiences, new use case, new verticals

Distributed functionality

5G

Longer latency
Big data/aggregated value
Content/storage/AI/processing



5G low latency
Customized/local value
Content/storage/AI/processing

Compute, vision, sensing
AI powered use cases
Internal AI optimizations

On-premise control for ultra-low latency

On-device intelligence assisted by cloud

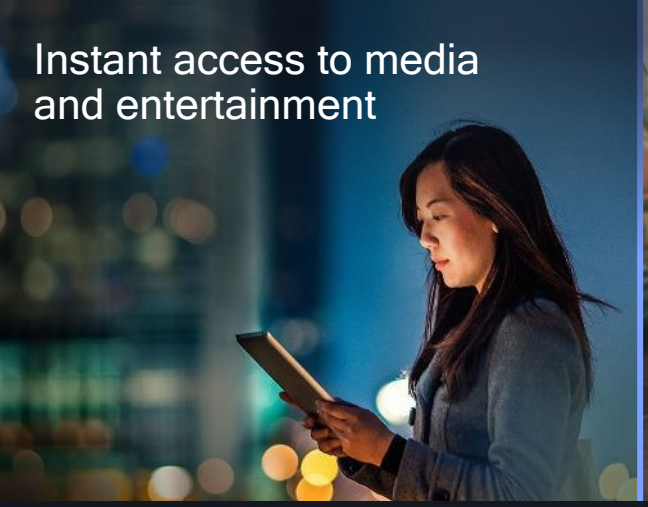
Distributed processing, like boundless XR

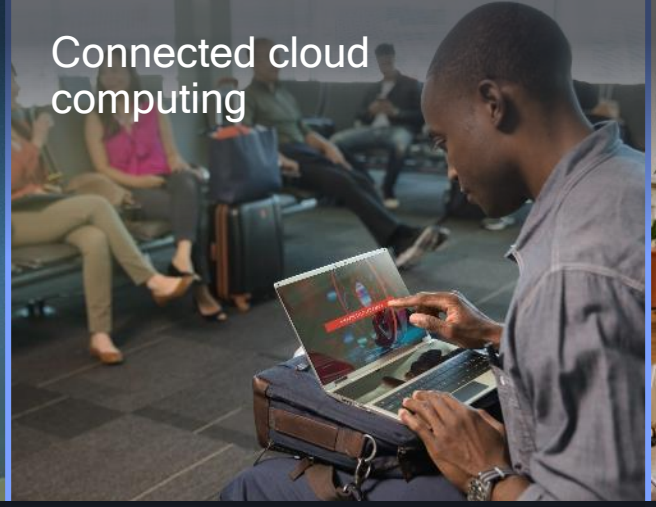
New services

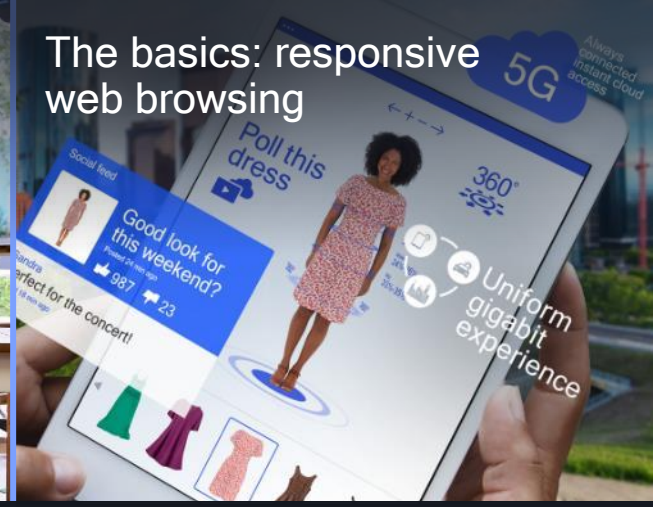
Cloud computing, storage, instant access

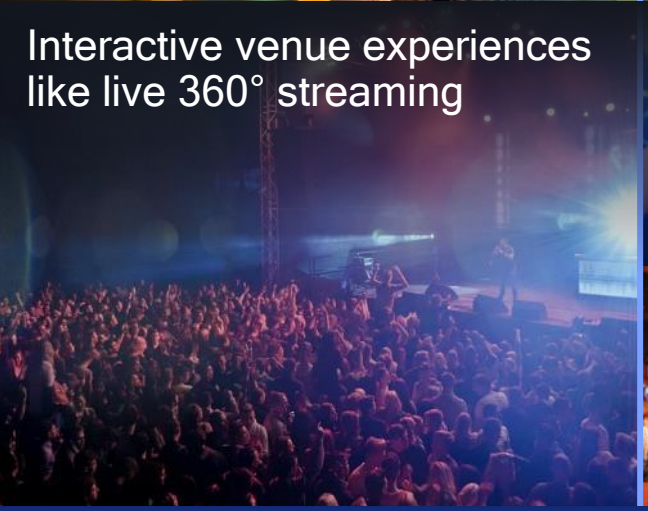
Low-latency gaming

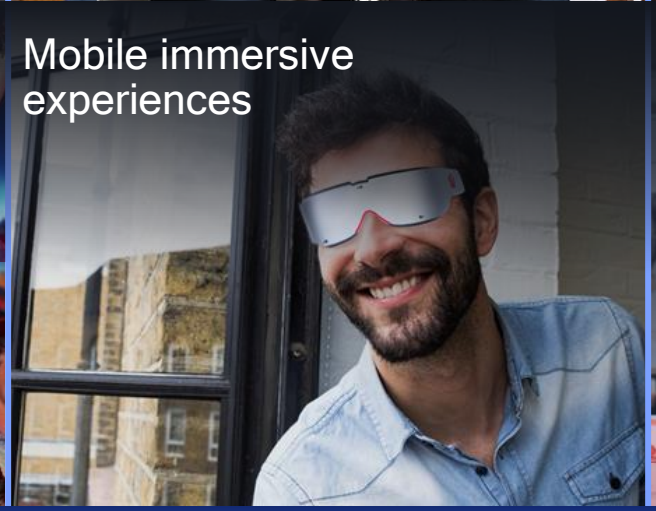
Real time assisted services like voice UI











5G

- Fiber-like data speeds
- Low latency
- Uniform performance
- Massive capacity

+

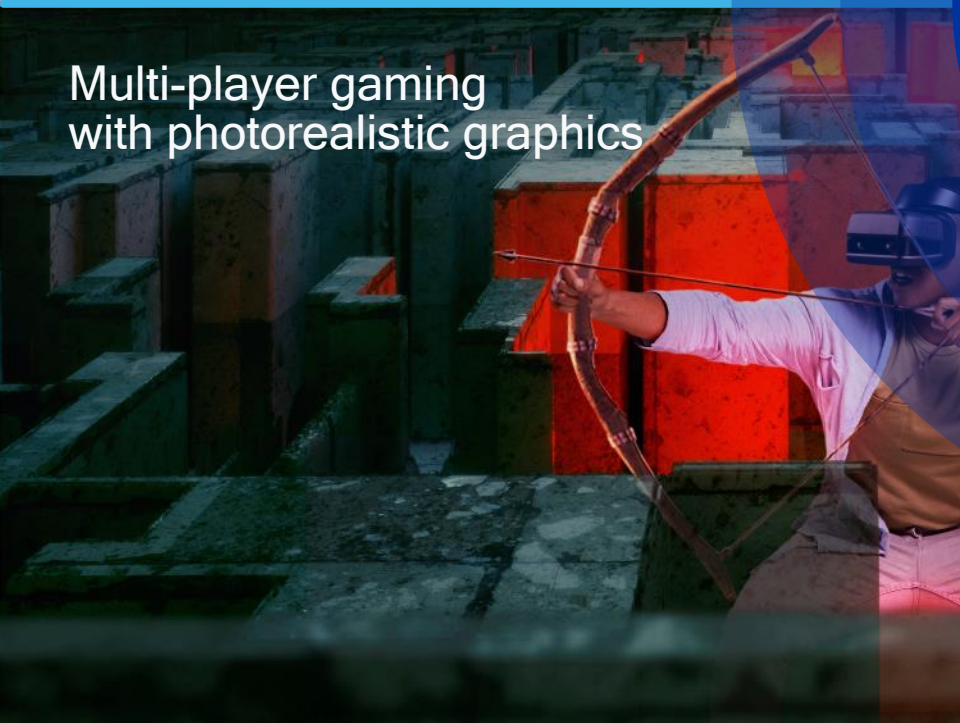
- Content/control closer to user
- Realization of low latency
- Customized local value
- Augment on-device processing

=

Enhanced and entirely new experiences

Real-time interactive collaboration

Shopping like never before



Multi-player gaming with photorealistic graphics

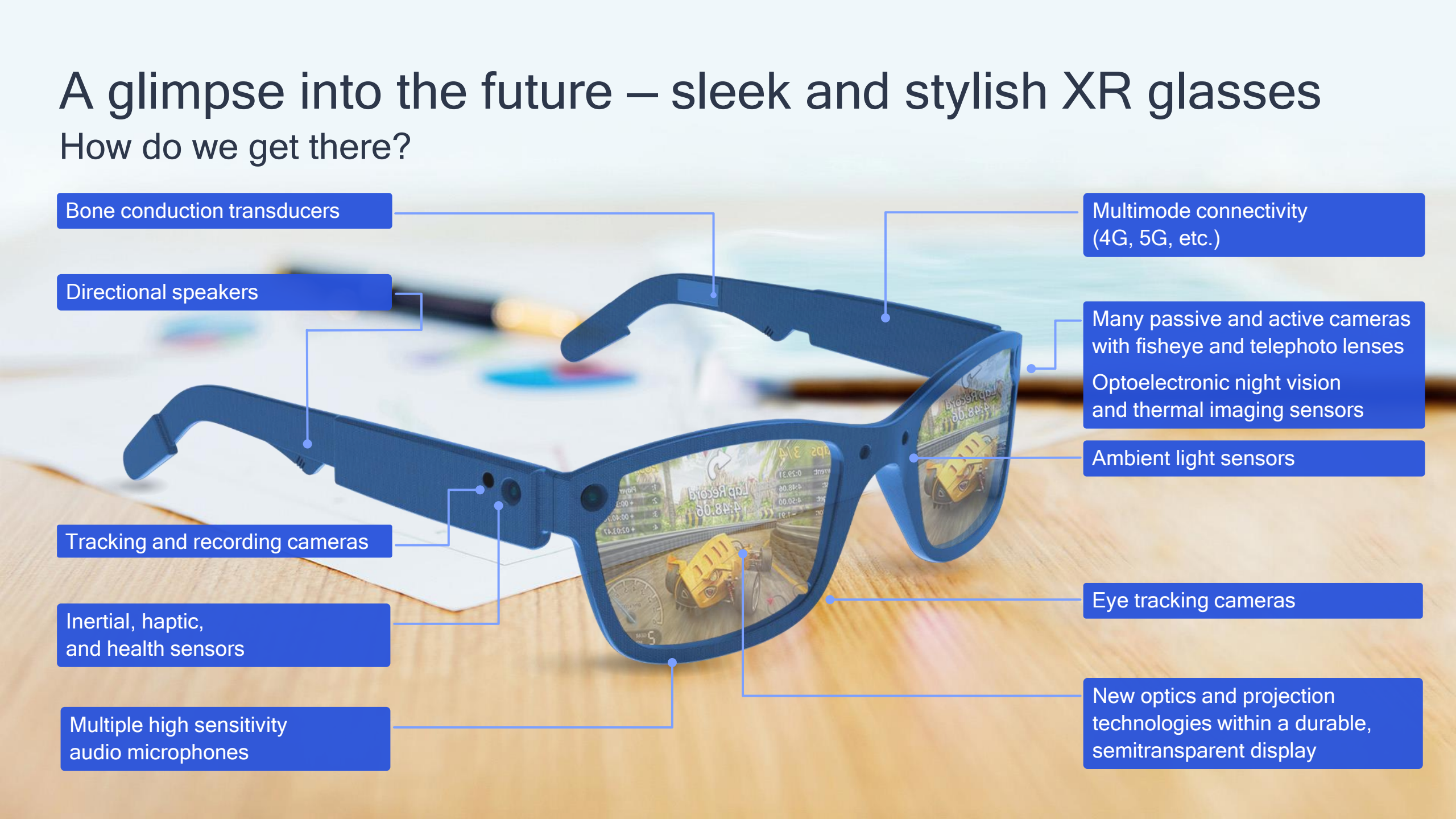
Our mobile office or living room, virtually anywhere

Boundless XR platform is ready for trials

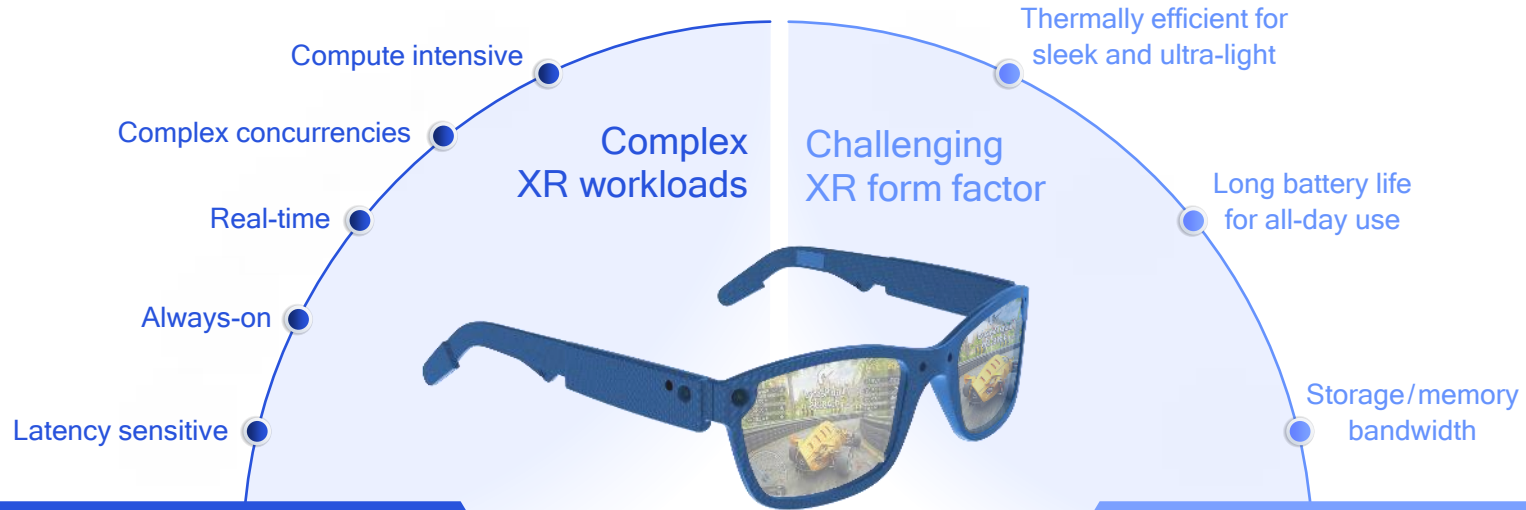
Rel-16 to 18 features will pave the way for low-power split-AR solution

A glimpse into the future – sleek and stylish XR glasses

How do we get there?



A new era in distributed processing



Essential on-device processing

Optimized under strict power, thermal, size constraints

Premium experiences today that continuously improve

Split rendering

5G

Low latency

High capacity

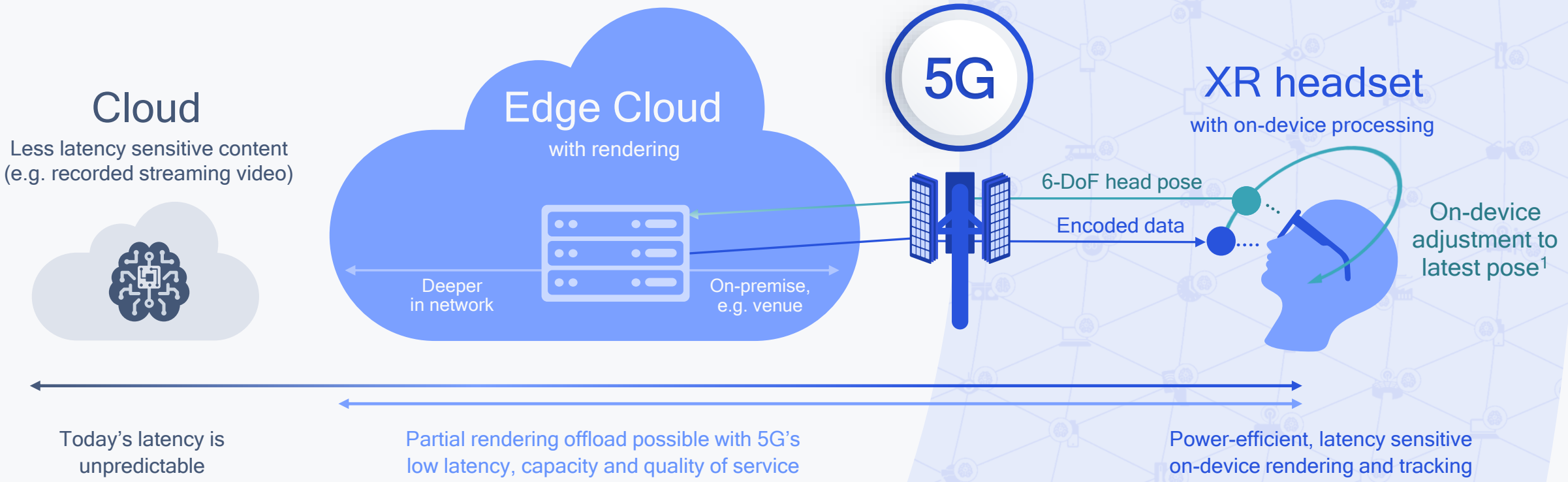
Reliable link

Augment by edge cloud processing

Significant higher power envelope–beyond PC class

Augment on-device rendering with edge cloud rendering

Augment on-device processing for boundless photorealistic mobile XR



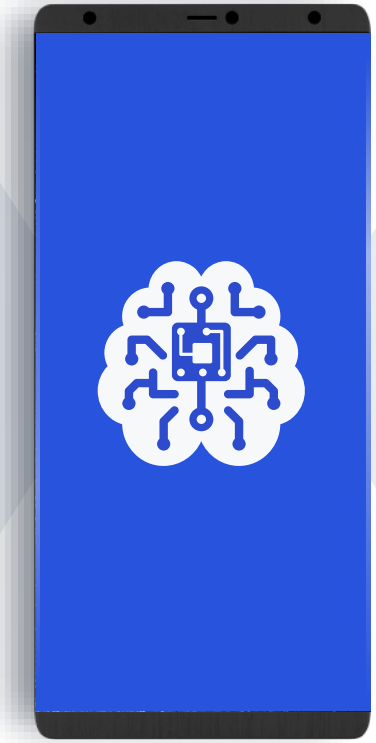
1. Asynchronous time warp reduces Motion to Photon (MTP) latency by using on-device processing based on the latest available pose. MTP below 20 ms generally avoids discomfort – has to be processed on the device

AI offers enhanced experiences and new capabilities for smartphones

True personal assistance

Extended battery life

Enhanced connectivity



Superior photography

Natural user interfaces

Enhanced security

A new development paradigm where things repeatedly improve

Voice is the transformative user interface (UI) we've been waiting for

Designed to be

Always-on

Conversational

Personal

Private

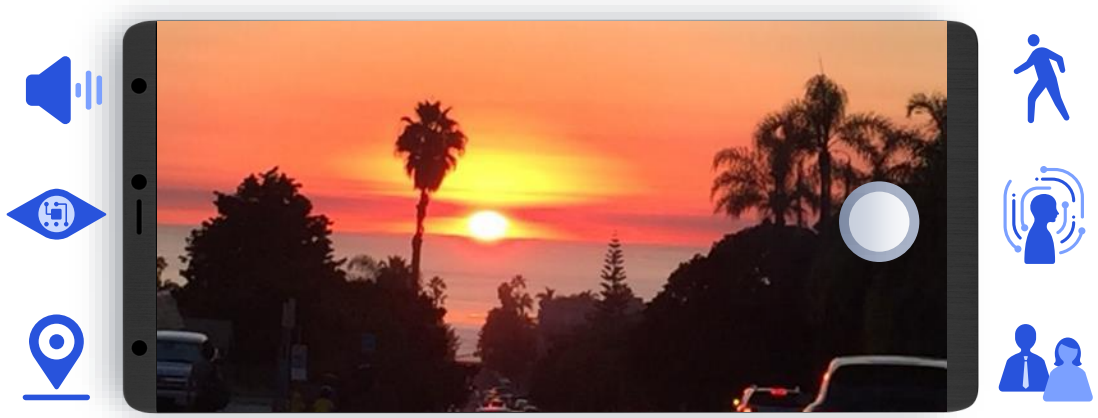
Critical to create a true virtual assistant

A context aware and personalized "digital me" sitting on the device

A true personal assistant is responsive and proactive

Responsive

Based on contextual analysis and prompting
(e.g. finding memories)



"Remember the time I was strolling with my son after the party at La Jolla beach?"

"Yes I do, here is a picture you took of the sunset. Should I share it with your family group on WeChat?"

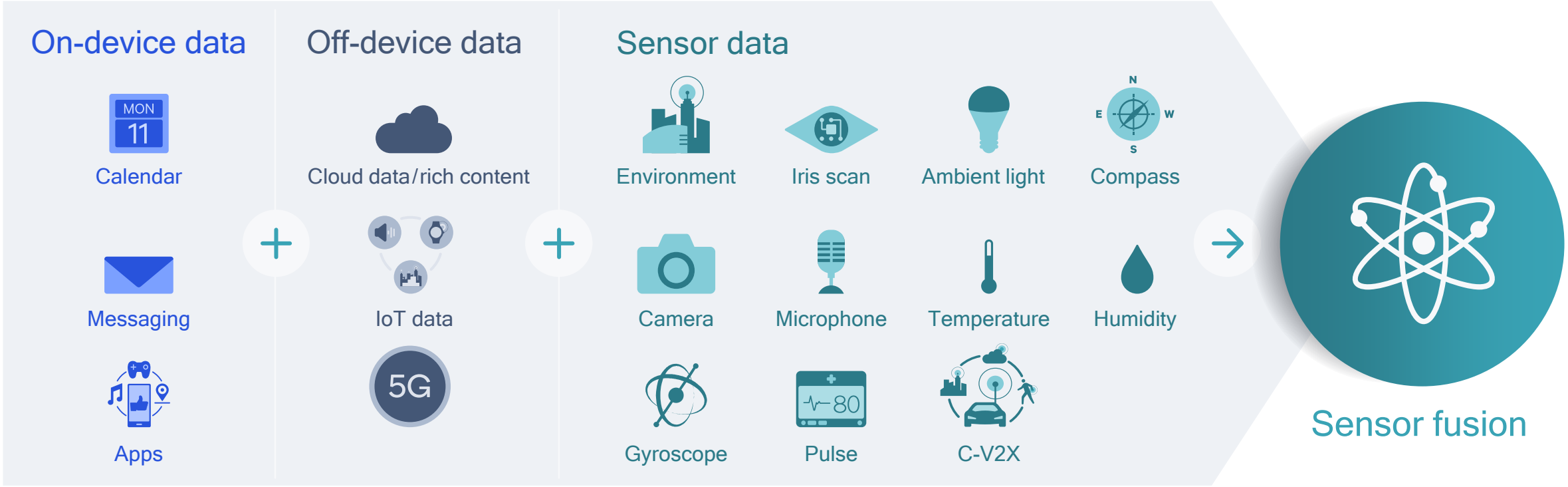
Proactive

Based on contextual analysis without prompting
(e.g. automatically sharing memories)

"I noticed that you are tired and stressed, I'm turning on the Rocky III soundtrack and navigating you to the gym for a workout and sauna."

"This music gets my blood going and a workout and sauna will help me relieve stress."

Contextual intelligence required for personalization

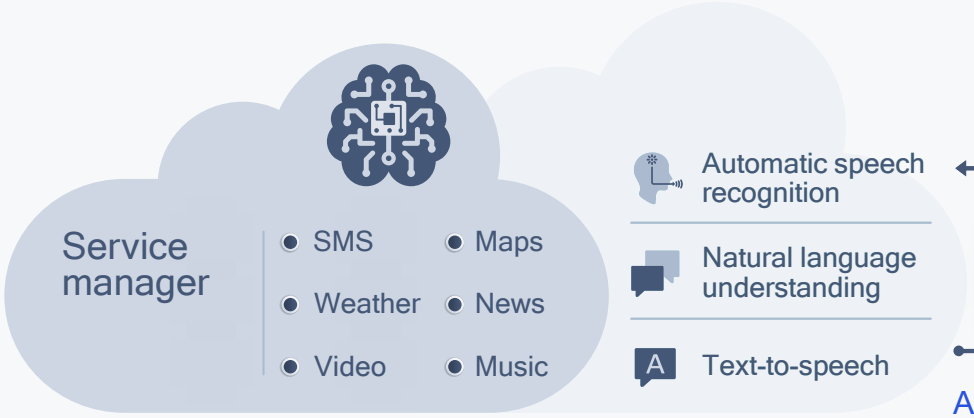


The local fusion of many types of sensors and personal information

On-device AI creates a true personal virtual assistant

Cloud-centric (today)

Less privacy and personalization



Automatic speech recognition

Natural language understanding

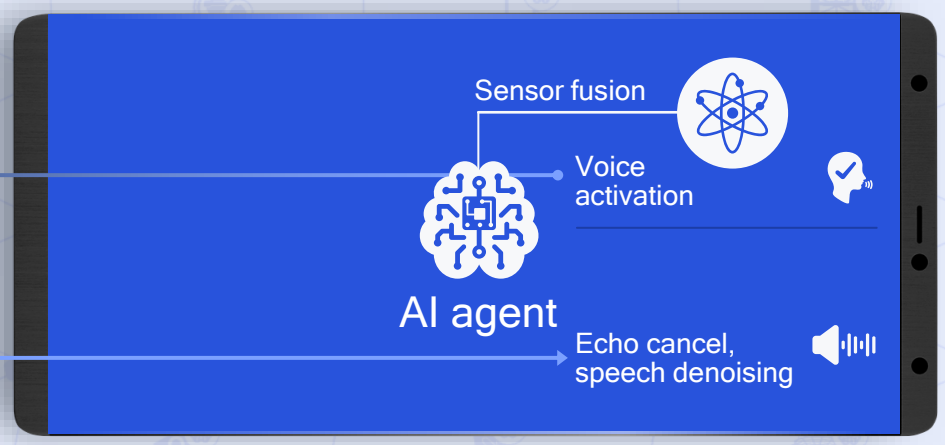
Text-to-speech

5G

Access to rich content

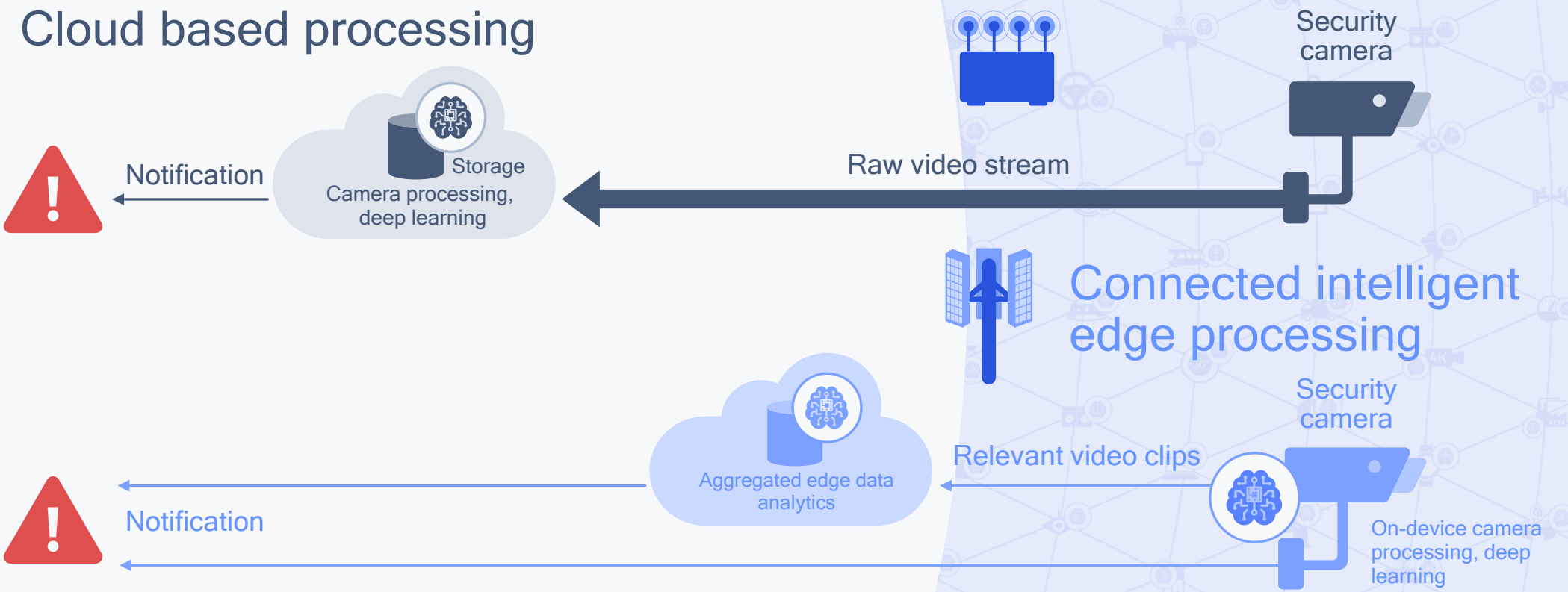
On-device (future)

Continuously learns personal knowledge and acts intuitively with immediacy



Both ends are needed – but AI functions gradually moving on to the device

All data should not go to the central cloud—especially private



Lower latency, privacy, on-device security, but also less bandwidth

AI in here. And everywhere.

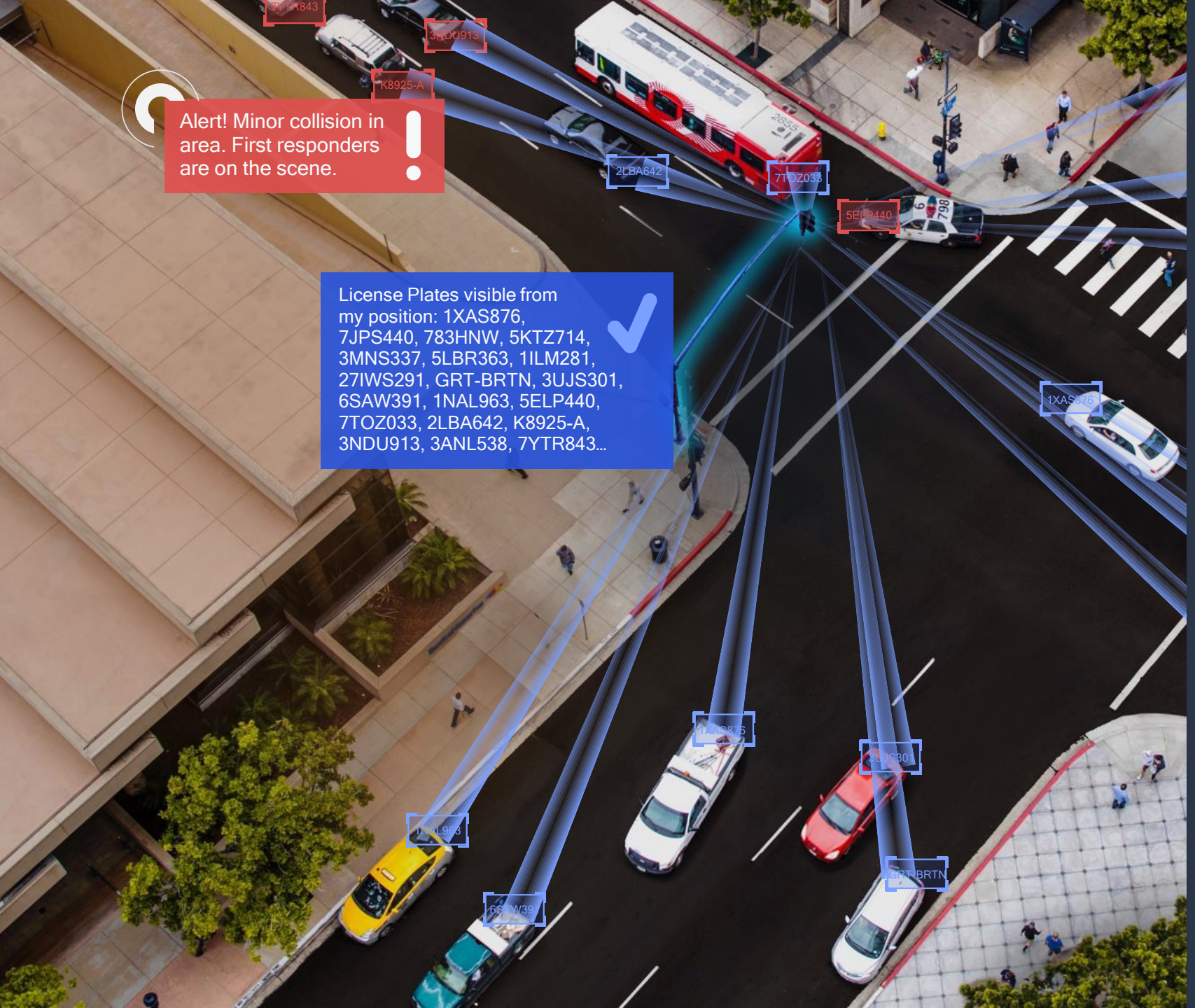
Advanced camera processing, powerful machine learning, and computer vision at the edge will allow new applications that push the boundaries of our connected world.

Intelligent cameras monitor and track what's inside.

Local at the edge processing means low latency and max efficiency.

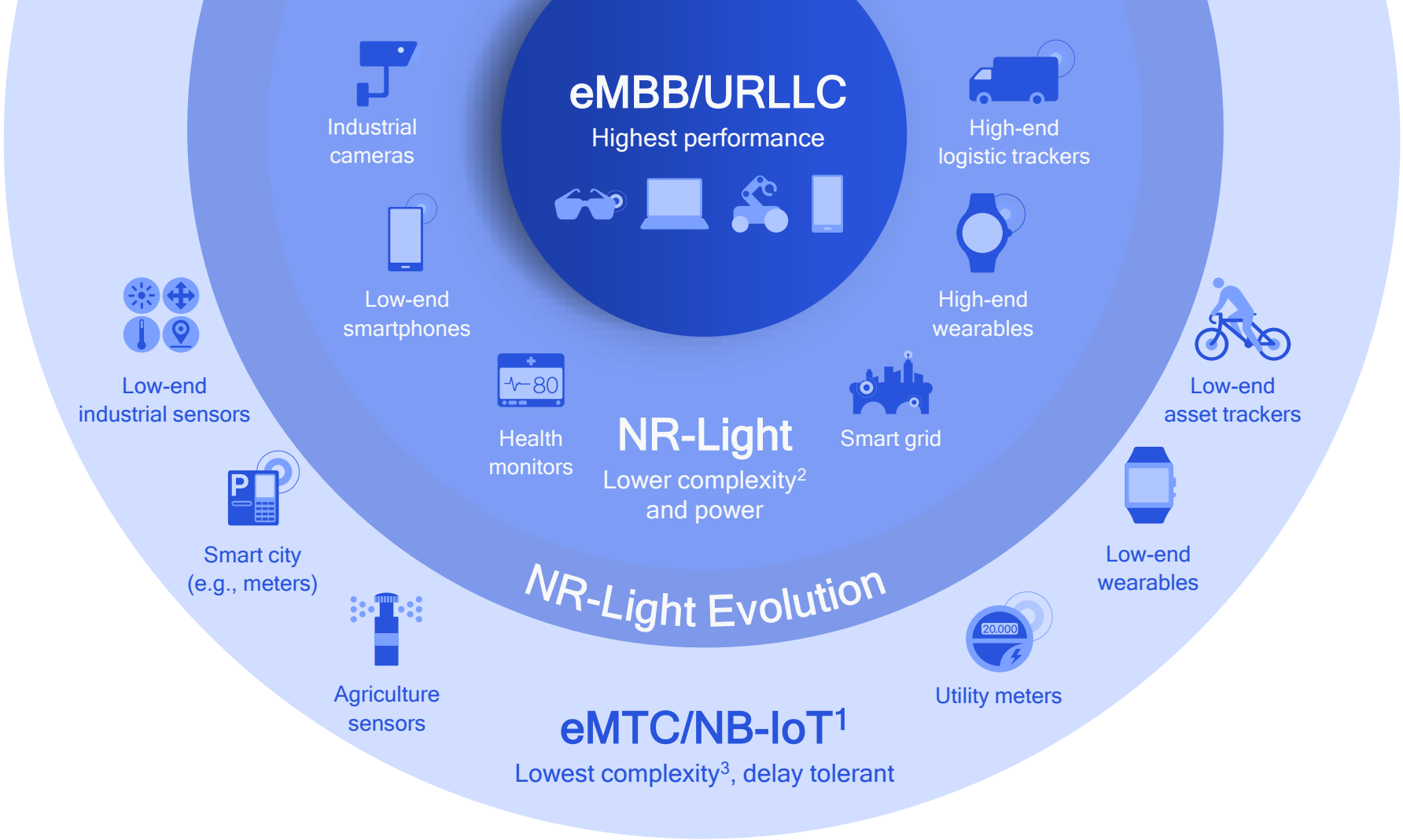
Neural networks watch and learn your preferences.

Ubiquitous connectivity helps you avoid running out of critical food supplies



Intelligent Smart Cities

- Scene Classification
- License Plate Recognition
- Target Sound Detection



5G NR: A unified, scalable air interface allowing coexistence of a wide range of 5G device classes

1 Also including satellite access; 2 Data rate of 150 Mbps DL / 50 Mbps UL, latency of 10-30 ms, 10-3 to 10-5 reliability, coverage MCL of 143 dB; 3 Data rate of 1Mbps, MCL of 155.7 dB (eMTC) and 164 dB (NB-IoT)

5G IoT needs new solutions at the connected intelligent edge

Opportunities

Monetize and manage massive amount of low-complexity things

Manage new, more intelligent devices and new use cases

New business and deployment model, like private networks for the Industrial IoT

Challenges

Trust and security

Large number of ecosystem participants

Total cost of ownership

HW/SW, life-cycle support, upgradeability

Our solutions

Trust environment rooted in HW

Device-cloud integration at scale

Zero-touch device life-cycle management

Trusted chipset services for enterprise and IoT providers

Trusted services to securely connect and manage devices

Qualcomm® wireless edge services cloud platform

Unique device key embedded in chipset at factory

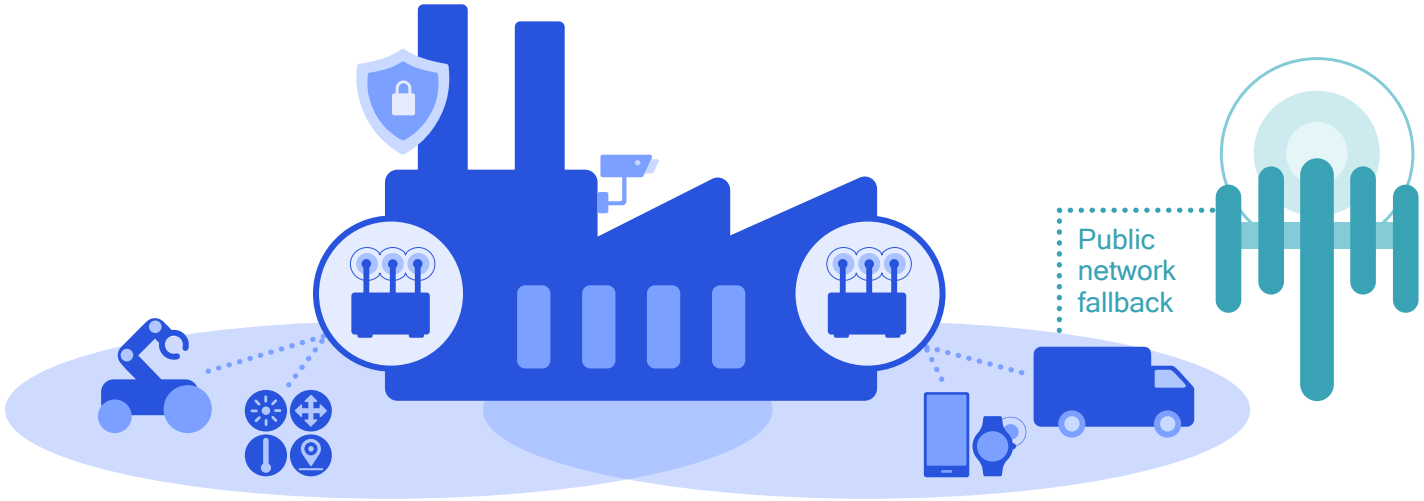
- Secure boot, debug
- Hardware root-of-trust
- Secure execution environment
- Runtime integrity checks

3rd party device management platform

Secure and trusted life cycle management

- Device attestation and connection integrity
- Device provisioning with plug-n-play on-boarding
- On-demand chipset upgrades and feature activation

5G private networks bring benefits to industry and enterprise



Private network[1]

Dedicated

Local network, dedicated resources, independently managed

Secure

Cellular grade security, sensitive data stays on-premise

Optimized

Tailored performance for local applications, e.g., low latency, QoS[2]

5G

Coverage, capacity, and mobility

Outdoor/indoor, high data speeds, seamless handovers, public network fallback

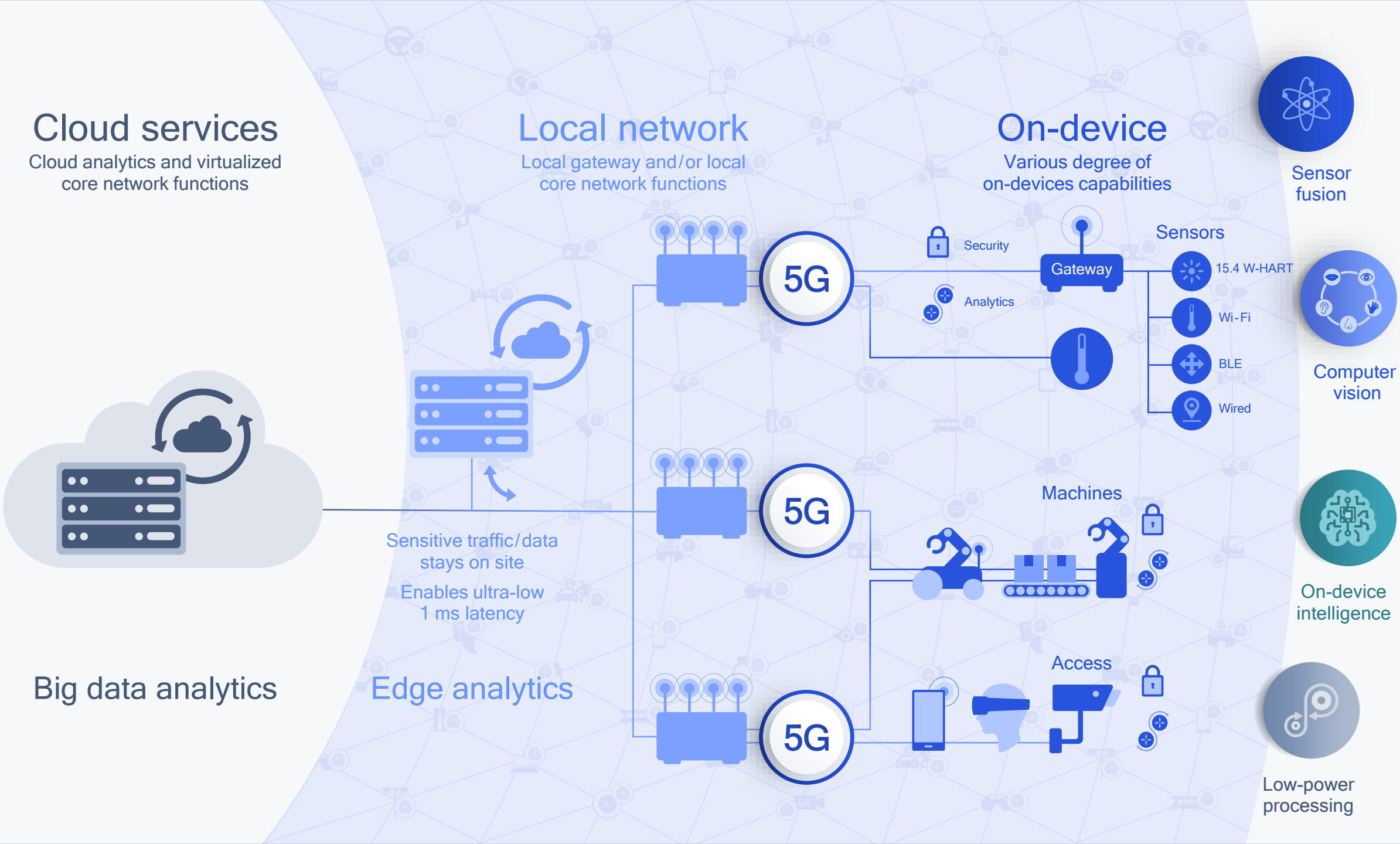
Reliability and precise timing

Industrial grade reliability, latency and synchronization (eURLLC[3] and TSN[4])

Interoperability

Global standard, vast ecosystem, future proof with rich 5G roadmap

1. Also referred to as non-public network (NPN); 2. Quality of service; 3. Enhanced ultra-reliable low-latency communication; 4 Time sensitive network


Cloud services
Cloud analytics and virtualized core network functions
Local network
Local gateway and/or local core network functions
On-device
Various degree of on-devices capabilities
Sensor fusion
Computer vision
On-device intelligence
Low-power processing
5G
5G
5G
Security
Analytics
Gateway
Sensors
15.4 W-HART
Wi-Fi
BLE
Wired
Machines
Access
Sensitive traffic/data stays on site
Enables ultra-low 1 ms latency
Big data analytics
Edge analytics


Qualcomm
Our technologies are already transforming the connected intelligent edge to realize the full potential of 5G
Learn more at www.qualcomm.com/research/5g
30°
80
20.000

Qualcomm

Thank you

Follow us on:

For more information, visit us at:

www.qualcomm.com & www.qualcomm.com/blog