June 2018

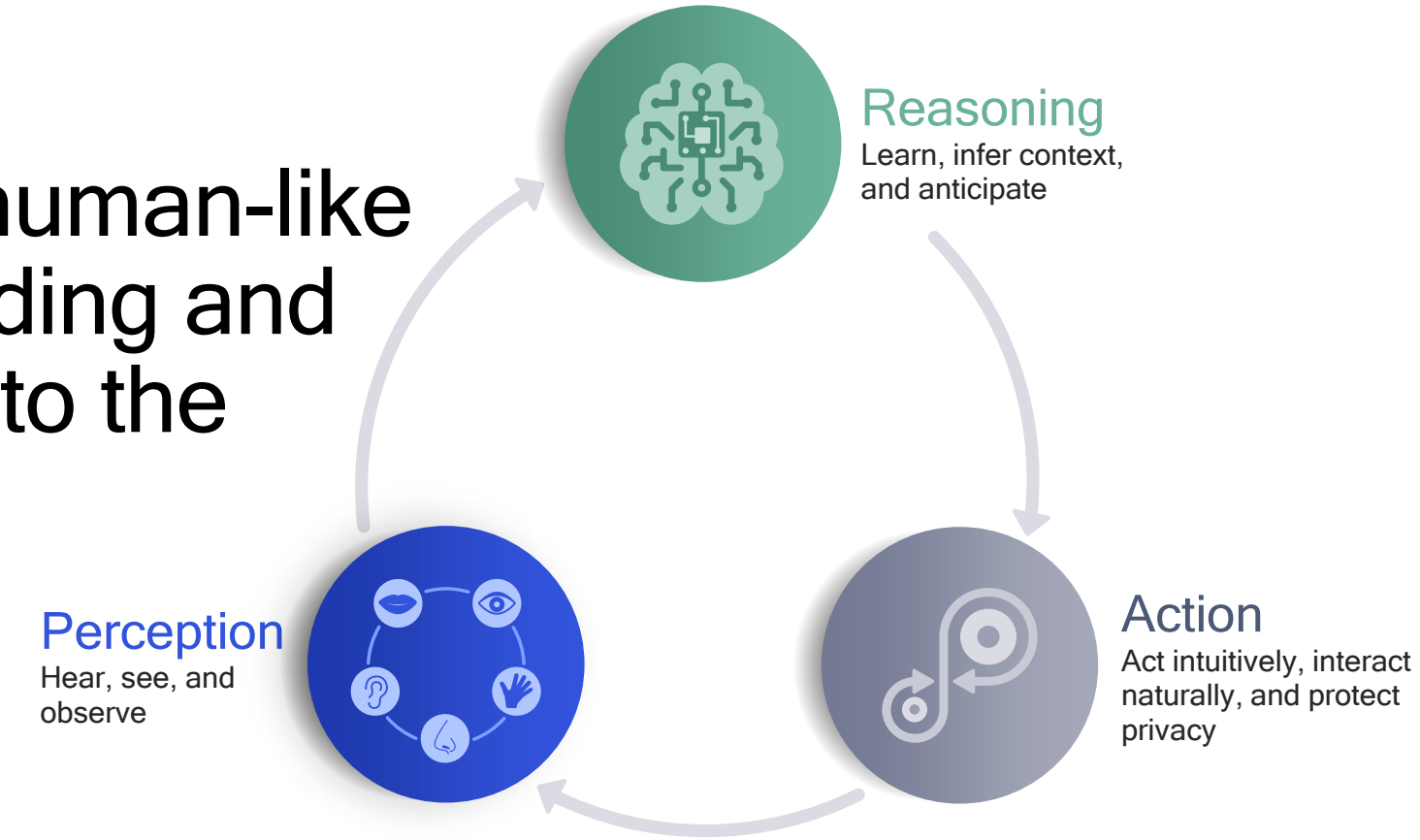
@qualcomm_tech

Qualcomm

Making an on-device personal assistant a reality

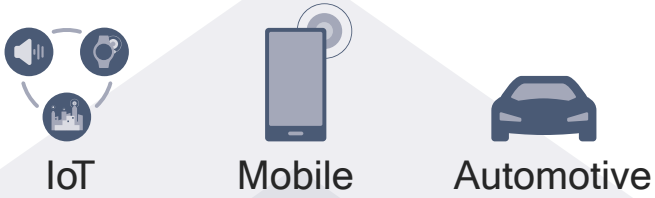
Qualcomm Technologies, Inc.

AI brings human-like understanding and behaviors to the machines

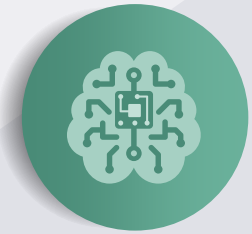


Advancing AI research to make on-device AI ubiquitous

A common platform is fundamental to scaling AI internally and across the industry



Perception
Object detection, speech recognition, contextual fusion



Reasoning
Scene understanding, language understanding, behavior prediction

Action
Reinforcement learning for decision making

Power efficiency
Model design, compression, quantization, activation, algorithms, and efficient hardware

Personalization
Continuous learning, model adaptation, and privacy-preserved distributed learning

Efficient learning
Robust learning through minimal data, unsupervised learning, and on-device learning

System architecture
Multi-task and multi-modal learning, sensor fusion, and cloud-edge systems

A true personal assistant

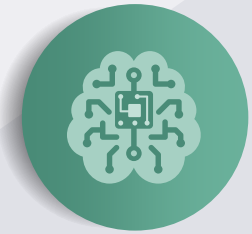
One of many use cases requiring a broad set of AI capabilities

IoT

Mobile

Automotive

Perception
Object detection, speech recognition, contextual fusion



Reasoning
Scene understanding, language understanding, behavior prediction

Action
Reinforcement learning for decision making

Power efficiency
Model design, compression, quantization, activation, algorithms, and efficient hardware

Personalization
Continuous learning, model adaptation, and privacy-preserved distributed learning

Efficient learning
Robust learning through minimal data, unsupervised learning, and on-device learning

System architecture
Multi-task and multi-modal learning, sensor fusion, and cloud-edge systems

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

Designed to be:

Always-on

Conversational

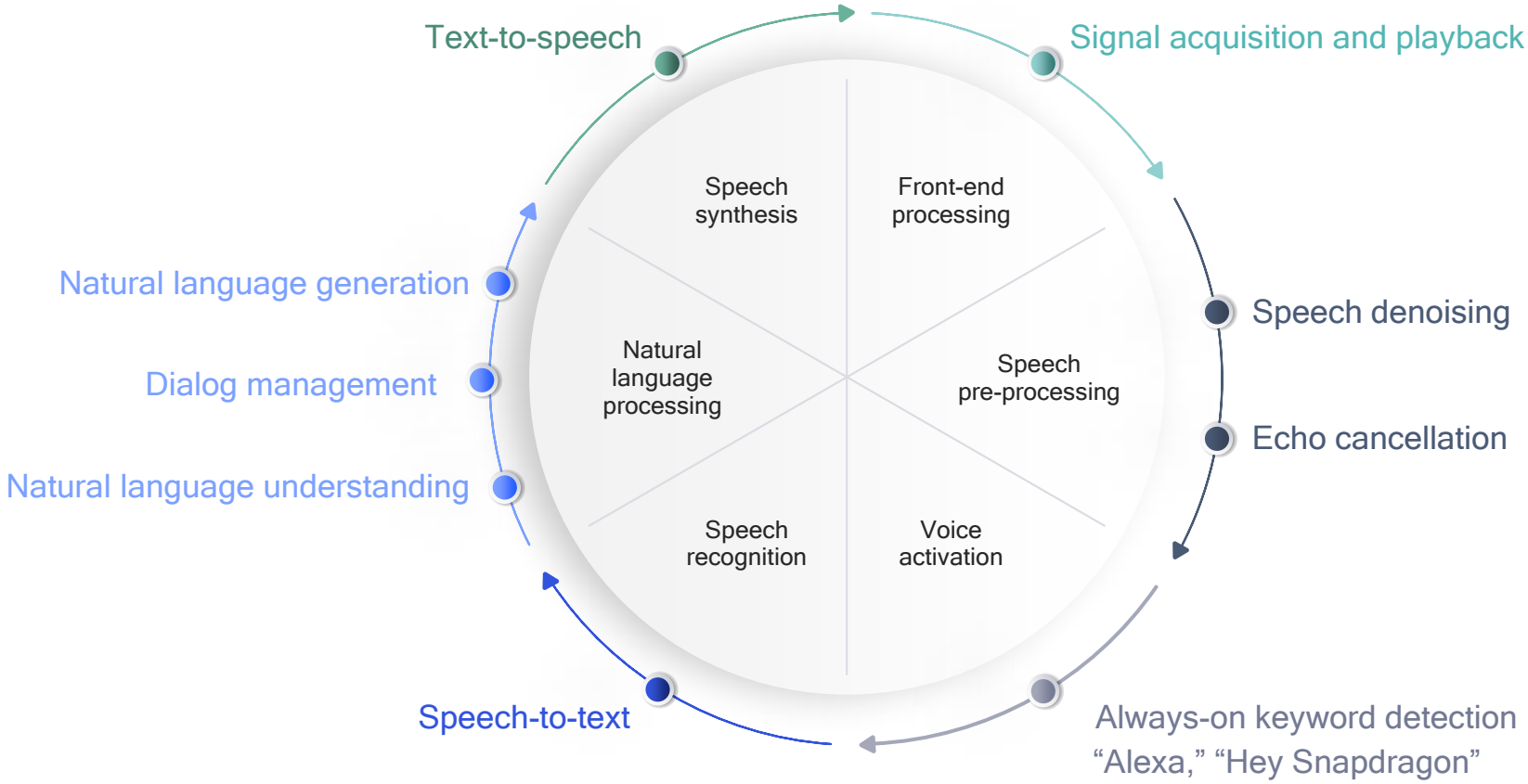
Personal

Private

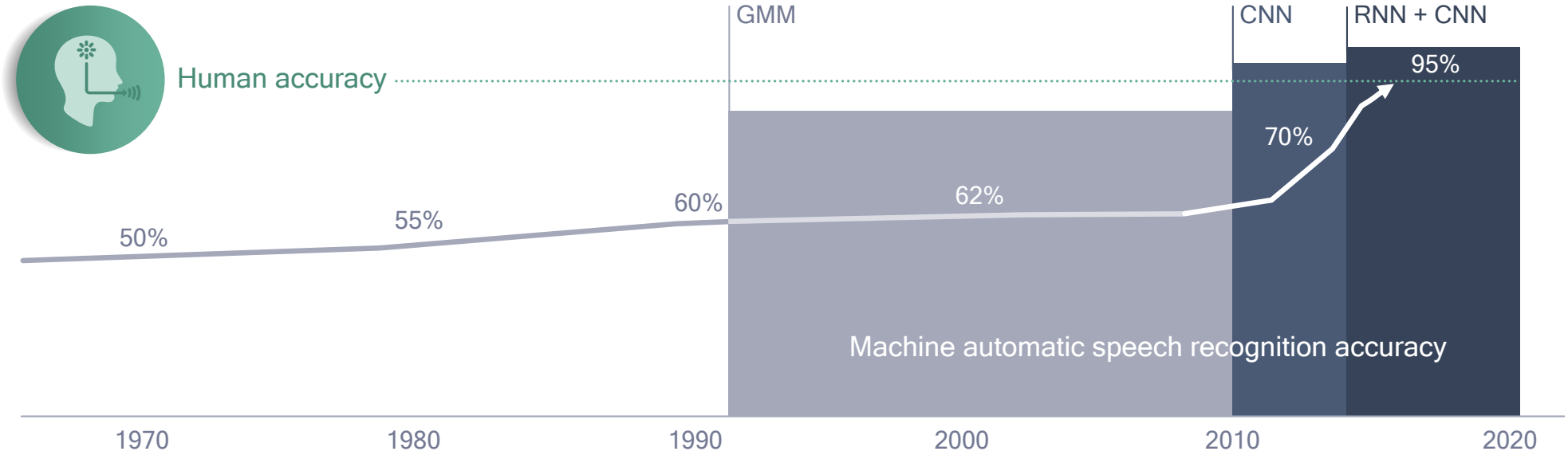
Critical to create a true virtual assistant

Voice UI components required for an end-to-end solution

Machine speech chain: listener and speaker



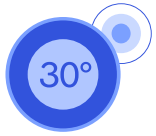
Machine learning has ignited the voice UI revolution



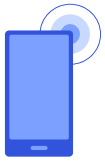
GMM: Gaussian Mixture Model, CNN: Convolutional Neural Network, RNN: Recurrent Neural Network

"As speech recognition accuracy goes from say 95% to 99%, all of us in the room will go from barely using it today to using it all the time. Most people underestimate the difference between 95% and 99% accuracy—99% is a gamechanger. No one wants to wait 10 seconds for a response. Accuracy, followed by latency, are the two key metrics for a production speech system."
—Andrew Ng

Voice UI is proliferating across product categories



IoT



Smartphones and tablets

Headphones and headsets

Wearables

Smart speakers

TV

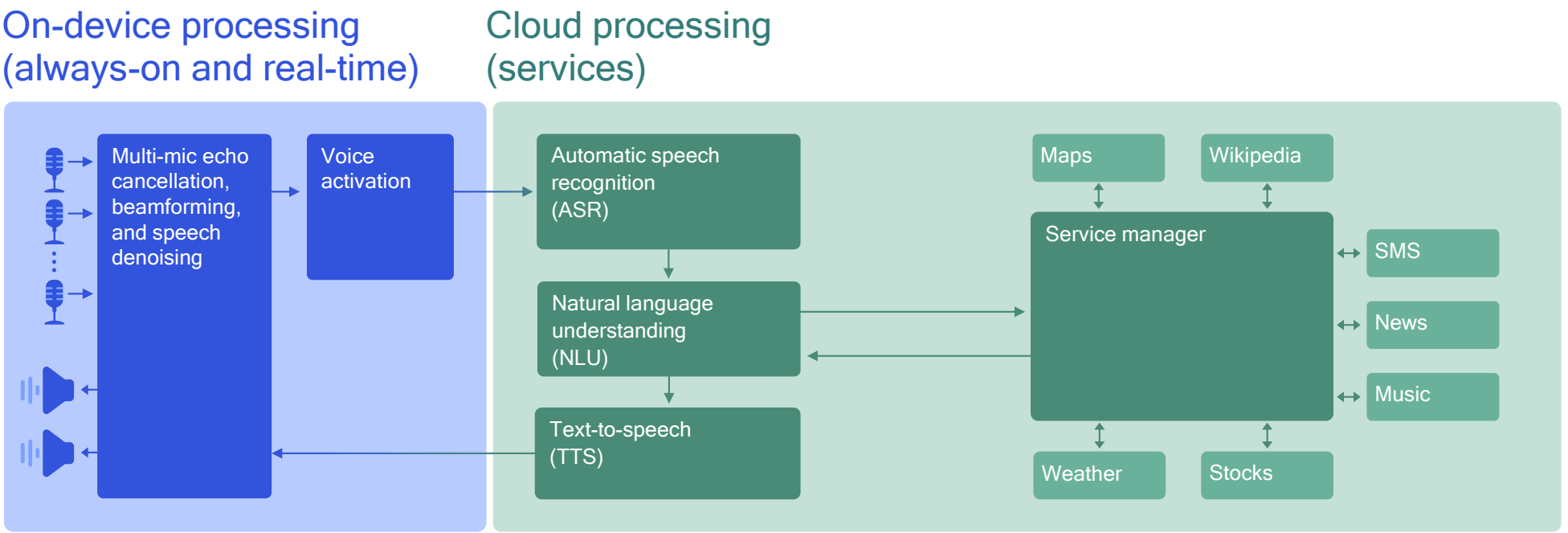
XR

In-car entertainment systems

PCs and laptops

Moving voice UI functionality to the end device

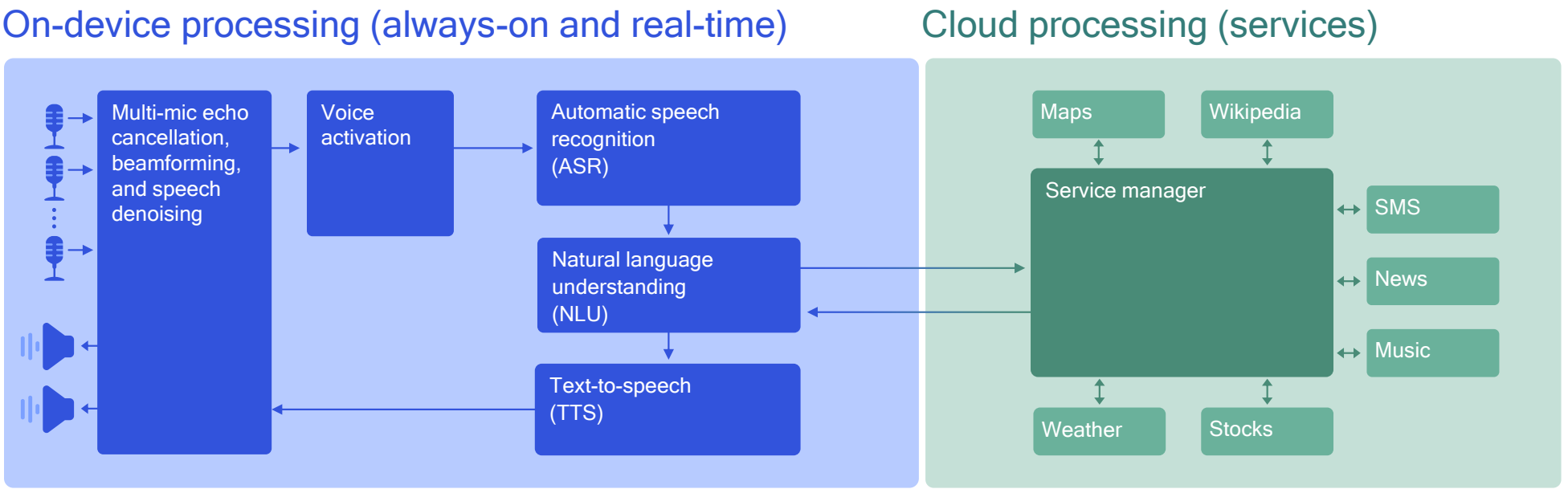
An end-to-end solution powered by machine learning



Cloud centric (today)

Moving voice UI functionality to the end device

An end-to-end solution powered by machine learning



On-device centric (future)

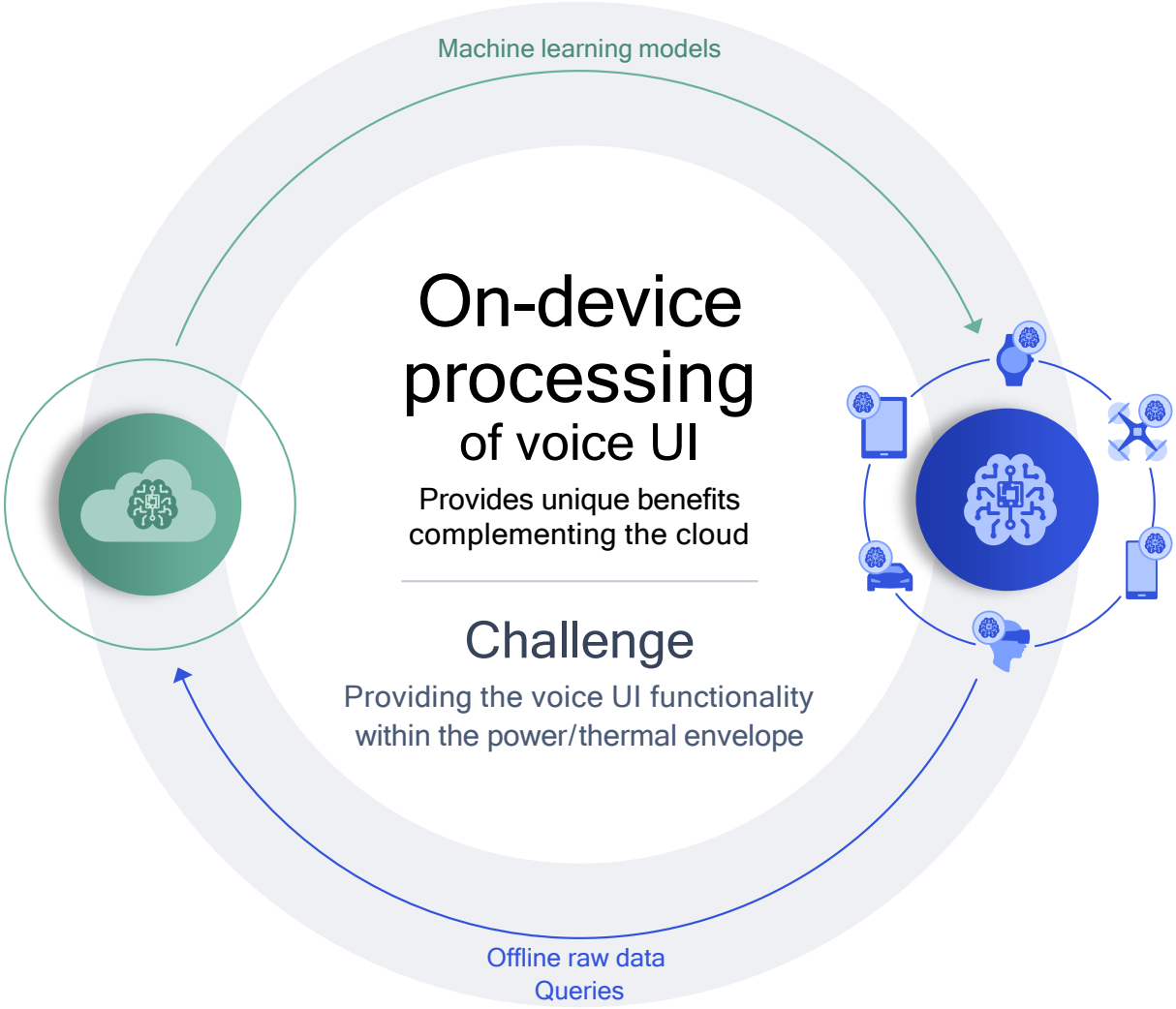
Cloud tasks

Complex voice fallback

Training and model update

Knowledge base

Services



On-device processing of voice UI

Provides unique benefits complementing the cloud

Challenge

Providing the voice UI functionality within the power/thermal envelope

On-device tasks

Automatic speech recognition

Natural language processing

Always-on audio cognition

On-device training

Benefits

Privacy

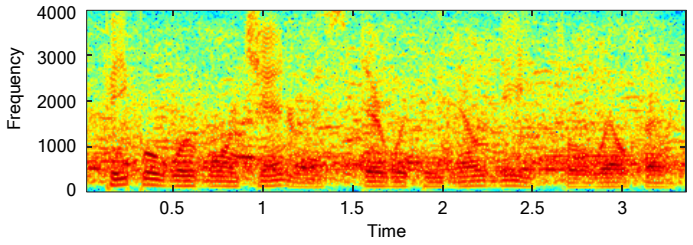
Instant response

Always-on

Device context

Speech denoising

- Single or multiple mics
- Applicable for
  - Two-way conversation
  - Voice/speaker recognition
  - Keyword spotting
- Deep learning (DL) significantly improves the performance over traditional methods
- Robust in challenging interference and noise scenarios

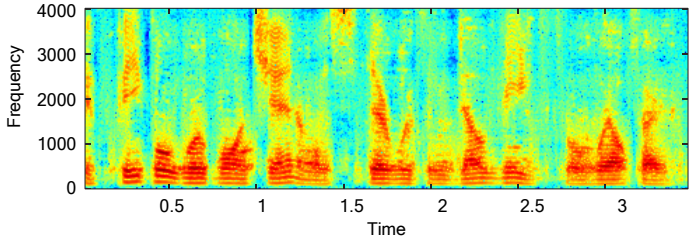


Noisy speech spectrogram

"If people were more generous, there would be no need for welfare"

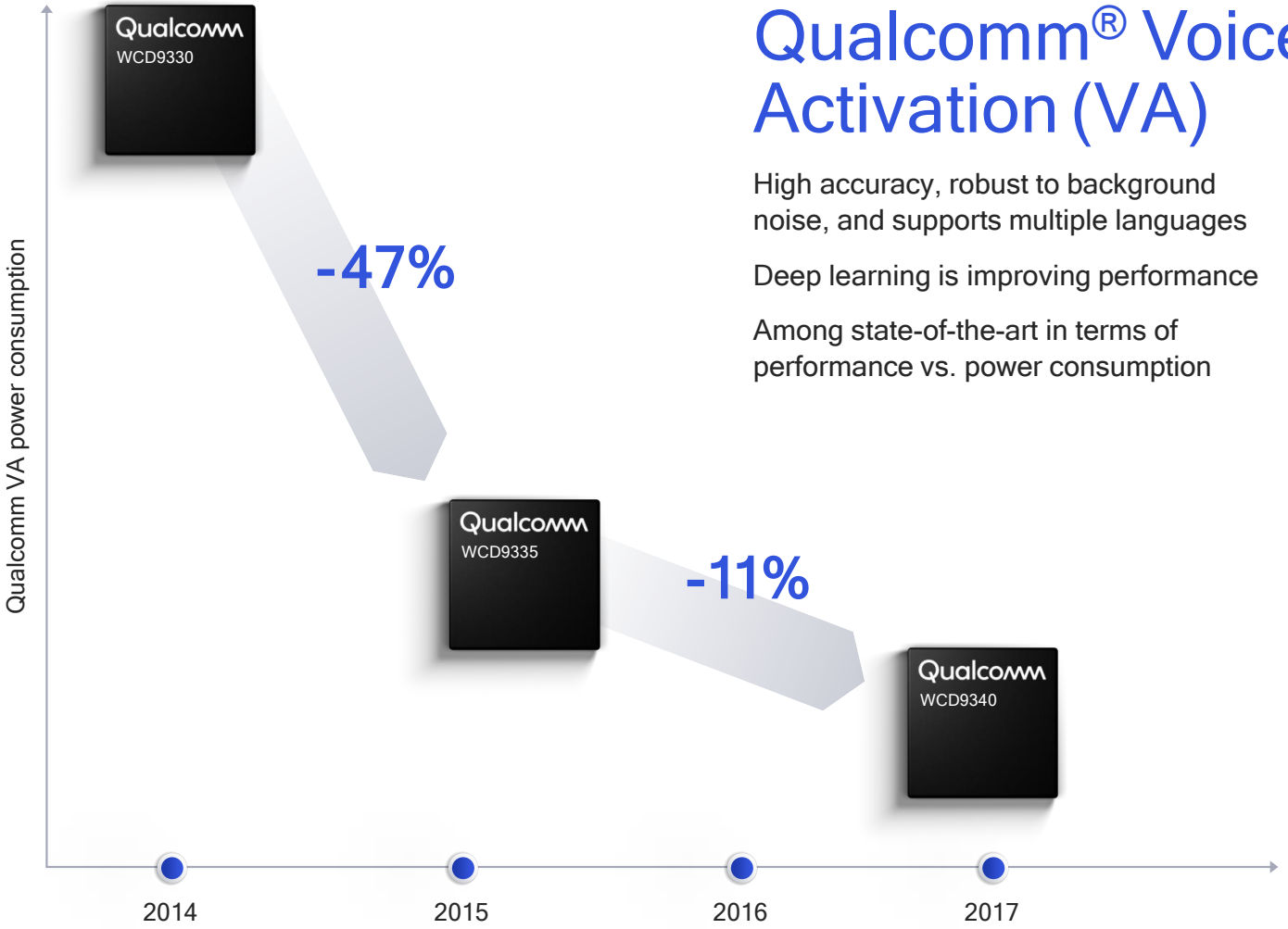
DL-based denoising

DL-based denoising model trained with extensive speech noise databases



Clean speech spectrogram

"If people were more generous, there would be no need for welfare"



Qualcomm® Voice Activation (VA)

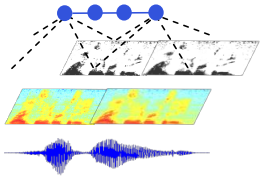
High accuracy, robust to background noise, and supports multiple languages

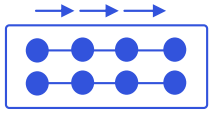
Deep learning is improving performance

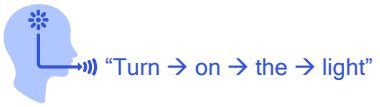
Among state-of-the-art in terms of performance vs. power consumption

Qualcomm Voice Activation supports:

- Amazon Alexa
- Baidu DUEROS
- Microsoft Cortana
- Google Assistant







Automatic speech recognition

Transcribe the audio to text

Deep learning gives state-of-the-art accuracy on a mobile device

---

Personalization–adaptation to individual accent and acoustic environment

On-device automatic speech recognition (ASR)

Acoustic features →

Reduce input audio to essential information

Acoustic model →

Deep learning converts input into linguistic units

Adapted to each user’s accent and environment

Language model

Uses context and language statistics for best utterance estimation

Adapted to each user’s speaking tendencies

↓

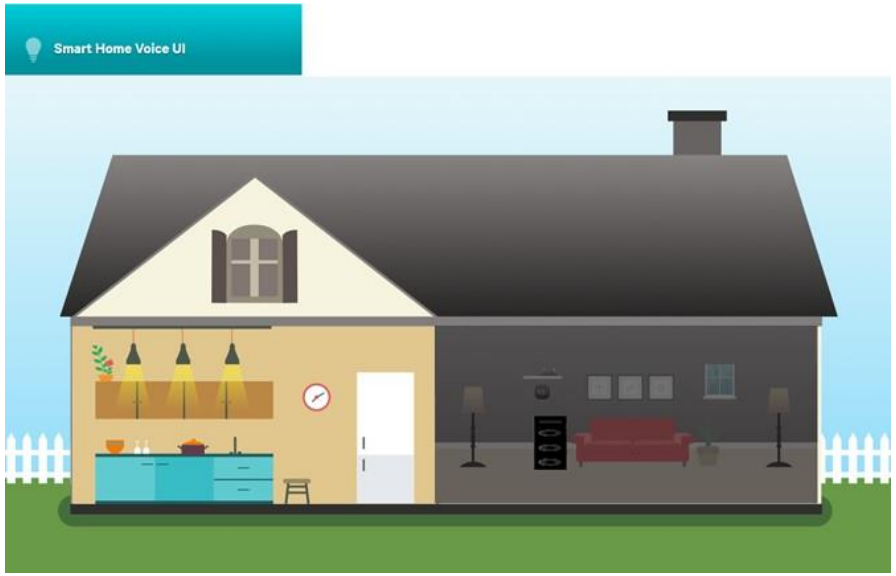
Natural language understanding (NLU) → User intention

Allows the same intention to be expressed in multiple ways

Adapted to each user’s intent expressions

An end-to-end on-device voice UI example for smart homes

Demo of automatic speech recognition and natural language understanding



Large command set

Turn on the living room lights
Click the kitchen lights off
Turn off all lights
Switch on the ceiling fan
Shut off the sprinklers
Start music
Pause song
Next track
Go back one
Play previous song
Turn speaker off
Increase temperature

Intent understanding

Turn on the kitchen light
Click kitchen light on
Switch on light in the kitchen
Turn the light on in the kitchen

NLU: These four phrases map to the same intent

99% on-device intent accuracy

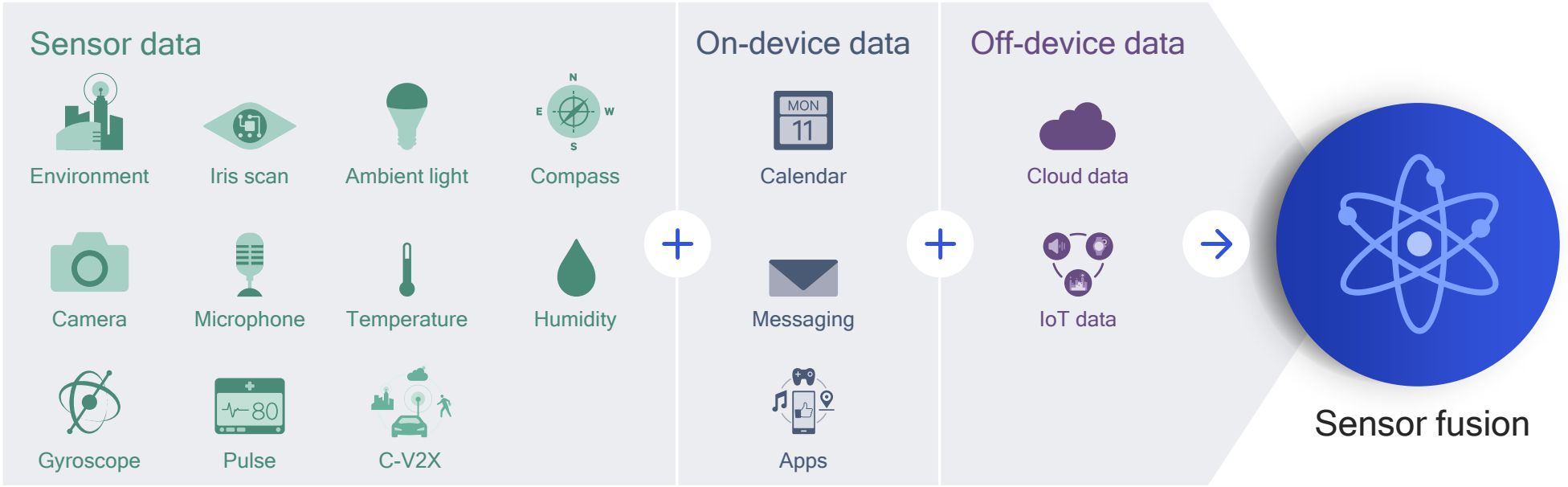
is achieved for domain specific command sets when adapted to accent and environmental condition

A true virtual assistant

A “digital me” sitting on the device: context aware and personalized

Contextual intelligence is required for personalization

The fusion of many types of sensors and personal information



Low power sensing, processing, and connectivity

- Efficient, heterogeneous architectures
- Sensor fusion and machine learning
- Integrated, always-on data capturing
- Low-energy wireless technologies (e.g. BT-LE, 5G NR IoT)

Creating personalized memories

Sound analysis

Talking with my son at sunset in La Jolla

Visual analysis

A sunset over the ocean in La Jolla

GPS location

La Jolla, California

Activity analysis

Strolling on the beach at sunset in La Jolla talking with my son

Live sentiment analysis

Strolling on the beach at sunset in La Jolla talking with my son and laughing

History, number of people, identity

After the party, strolling on the beach at sunset in La Jolla talking with my son and laughing

Essential for a true virtual assistant

A true personal assistant is responsive and proactive

Responsive

Decision-making and conversation 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

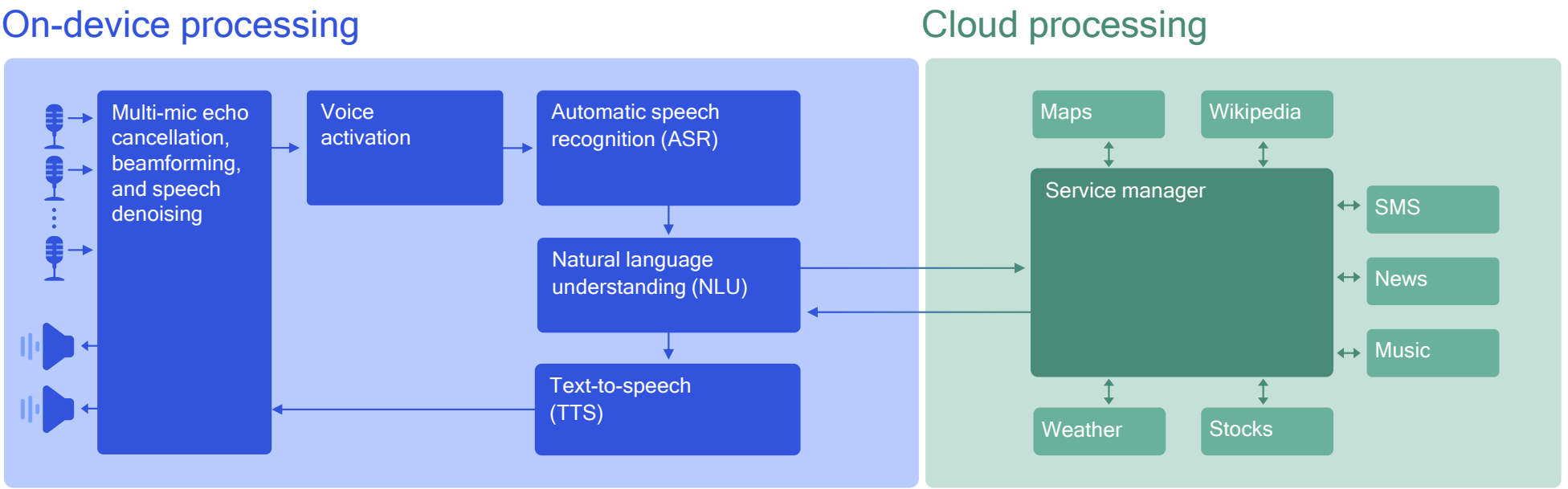
Decision-making and conversation based on contextual analysis without prompting (e.g. automatically sharing memories)

"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."

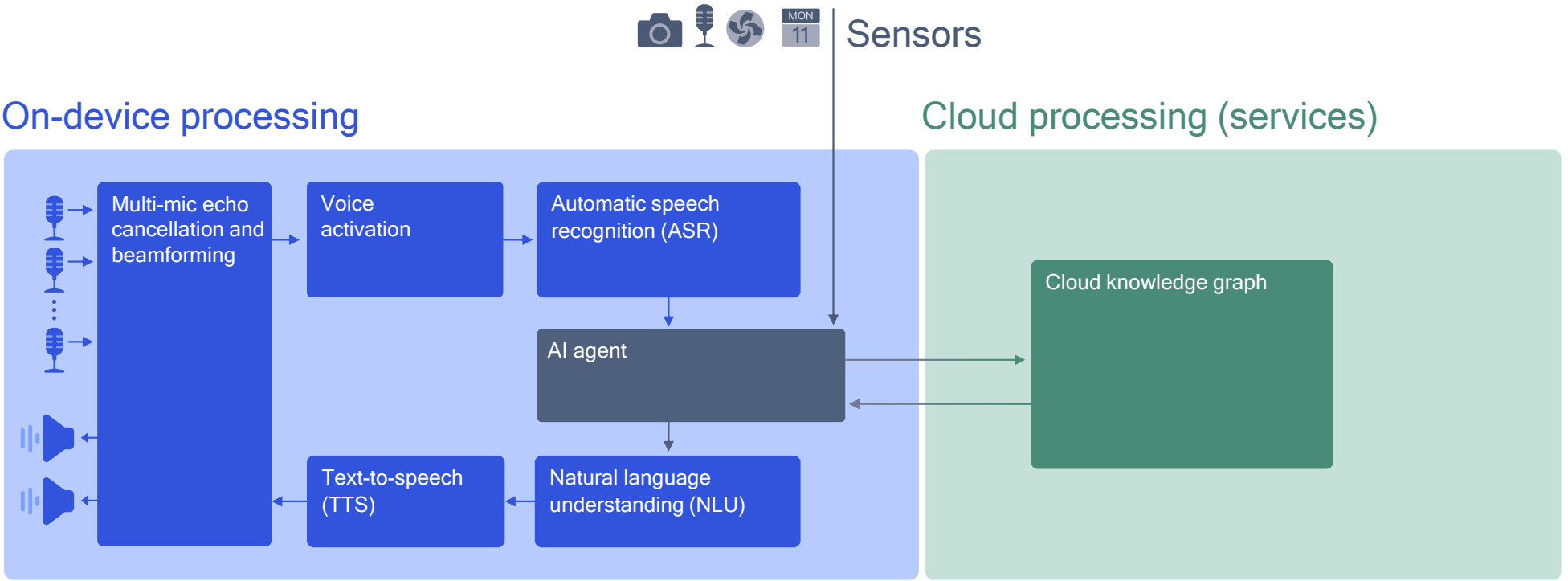
The first step to an on-device virtual assistant

Enabling on-device voice UI



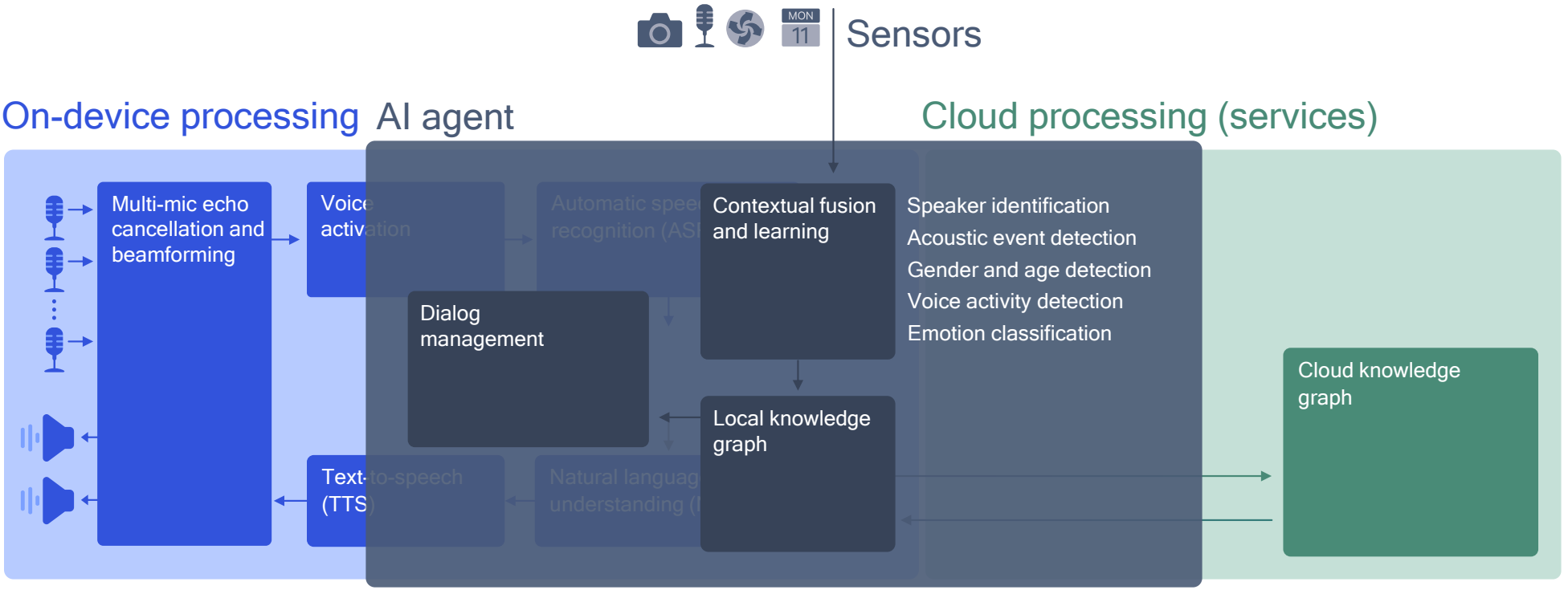
Adding an “AI agent” to create a true virtual assistant

The on-device AI agent continuously learns personal knowledge and acts intuitively



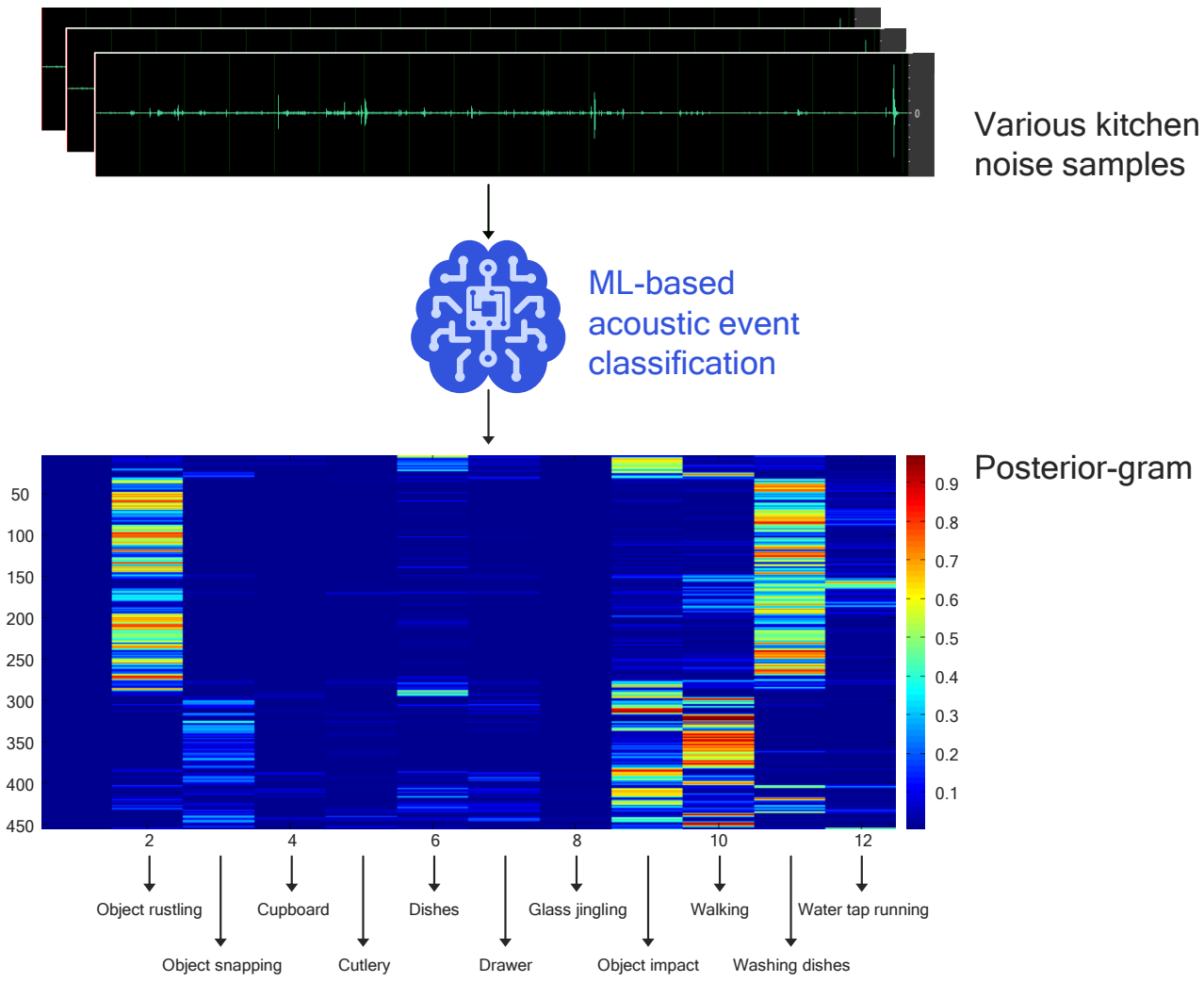
Adding an “AI agent” to create a true virtual assistant

Contextualization allows personalization at acoustic, intent, and behavior levels



Acoustic event detection

- ML techniques are used to
  - Classify acoustic signals into a set of predefined events
  - Infer acoustic environment
- Low power, always-on



Qualcomm

We are advancing AI research to make on-device AI ubiquitous

We are creating AI platform innovations that are fundamental to scaling AI across the industry

We provide the low-power end-to-end on-device solution for a true personal assistant

Qualcomm

Thank you!

Follow us on:

For more information, visit us at:

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