

## Aria Gen 2: An Advanced Research Device for Egocentric AI Research

By The Project Aria Team at Meta

### Abstract

Egocentric perception from wearable multimodal devices is crucial for contextual and embodied AI. Project Aria, developed by Meta Reality Labs Research, is a platform to advance egocentric sensing and real-world AI. The [first-generation Aria glasses](#), launched in 2020, established a foundation for large-scale, synchronized egocentric data capture. With over 1,000 units deployed globally, Aria is a widely adopted device for egocentric research, supporting diverse datasets in perception, interaction, and environmental modeling.

This paper introduces Aria Gen 2, a new generation of all-day wearable smart glasses that expand the boundaries of multimodal egocentric sensing. The device combines improved ergonomics and extended battery life with enhanced sensing fidelity and on-device computation for low-latency machine perception. With integrated real-time pipelines for six-degree-of-freedom (6-DoF) tracking, articulated hand pose estimation, and eye-gaze inference, Aria Gen 2 supports both recording and live streaming of contextual perception signals. These capabilities reduce the friction of dataset creation and open new opportunities for interactive, embodied, and context-aware AI systems. Together with its accompanying SDK and software tools, Aria Gen 2 provides a scalable, reproducible research framework for the next generation of egocentric perception studies. By uniting advances in sensing, compute, and usability, this platform aims to bridge the gap between machine perception research and deployable intelligent agents—paving the way toward truly personalized, contextually aware augmented reality experiences.

### I Introduction

Smart glasses with multimodal sensor suites promise a future paradigm of computing that can interact seamlessly with AI systems. Project Aria and the first-generation Aria glasses were introduced to the research community in 2020 as a step-change in all-day wearable sensing. The platform combined multi-sensor, hardware-synchronized egocentric recording

and streaming, enabling large-scale, naturalistic egocentric dataset development for the first time. Since its debut, Aria Gen 1 has become the world’s most widely adopted device for egocentric research, with over 290 academic and industrial partners operating more than 1,000 devices across 27 countries. Meta has released a suite of datasets derived from Project Aria that address the fundamental problems of understanding people within their environments, modeling the environments themselves, and capturing how humans interact with and alter them. Together, the device and these datasets have generated hundreds citations across research areas including computer vision, contextual AI, robotics, augmented reality, human–computer interaction, and assistive technologies.

Announced in February 2025, **Aria Gen 2** introduces a new generation of egocentric research capability. The device extends all-day wearability with improved ergonomics and longer battery life, and adds specialized on-device computation for low-latency machine-perception tasks including 6-DoF pose tracking, articulated hand keypoint estimation, and eye-gaze inference. These perception signals can be recorded or streamed to a host PC, reducing end-to-end dataset-creation friction and enabling new real-time applications such as robotic manipulation and embodied-AI experimentation. Enhanced audio I/O further supports interactive experience prototyping and speech-in-context research. Collectively, Aria Gen 2 represents a significant step toward real-time, contextually aware, and embodied AI systems that bridge perception research and deployable intelligent agents. We believe that this device, with its improved wearability and integrated machine perception capabilities, will pave the way for advancements in [contextual AI](#) and robotics research.



	Aria Gen 1 	Aria Gen 2 
Battery life	1-2 Hours	6-8 Hours
Sensors	Front-facing 8MP RGB Camera 2 Computer Vision Cameras Eye Tracking Cameras Dual 6-axis IMUs Barometer Magnetometer 7 Spatial Microphones Global Navigation Satellite System (GNSS)	Front-facing 12MP RGB Camera 4 Computer Vision Cameras <b>(wider device FOV, front-facing stereo pair, HDR)</b> Eye Tracking Cameras Dual 6-axis IMUs Barometer Magnetometer 7 Spatial Microphones Global Navigation Satellite System (GNSS) <b>Proximity Sensor</b> <b>Contact Microphone</b> <b>Photoplethysmography (PPG) health sensor</b> <b>Ambient Light Sensor</b>
Machine Perception	Offline using Machine Perception Services	Real-time on-device with custom Meta co-processor for core algorithms and offline using Machine Perception Services for additional algorithms
Interactivity	Interactivity via streaming to PC and using PC display or audio	Contact Microphone and Open Air Speakers enable direct interactivity with device
Streaming	Sensor data streamable to local PC via USB Cable or Wirelessly	Sensor data and on-device machine perception signals streamable to local PC via USB or wirelessly
Form Factor	Not Foldable	Foldable

Table 1. Overview of the high-level differences between Aria Gen 1 and Gen 2 devices.

This paper is organized as follows. Section II introduces the Gen 2 device, illustrating the hardware advancements and advanced on-device software capabilities. Section III presents the client SDK, which enables control of the Gen 2 devices. Finally, Section IV discusses the motivational research frontier unlocked by this device.

## II Device Overview

### Wearability and Design

Aria Gen 2 is designed with research usability and comfort in mind. The glasses feature a lightweight form factor (approximately 75 grams) and foldable arms, making them easy to wear for extended periods and portable for daily use. The design prioritizes all-day usability, supporting six to eight hours of continuous operation, which is essential for long-duration research studies in real-world environments. The open-ear, force-canceling speakers provide audio feedback without occluding environmental sounds, enabling natural interactions and user-in-the-loop prototyping.



There is a privacy switch on the glasses that prevents the device from capturing audio, video, or other sensor data when engaged. This tangible safeguard assures users and bystanders that no recording is occurring, even if other controls or software settings are active. The outward-facing privacy LED signals to bystanders when the device is actively capturing content, such as photos, videos, or livestreams. This LED is located on the front of the glasses for visibility to those around the wearer.

## New Sensors Unlocking Research Opportunities

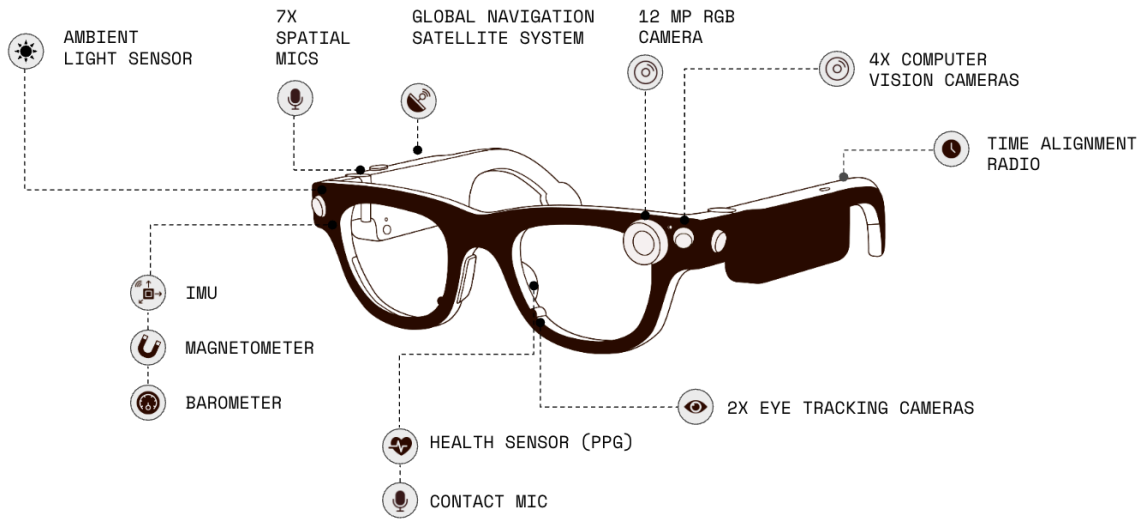


Figure 2a. Illustration of the sensor suite of the Aria Gen 2 device

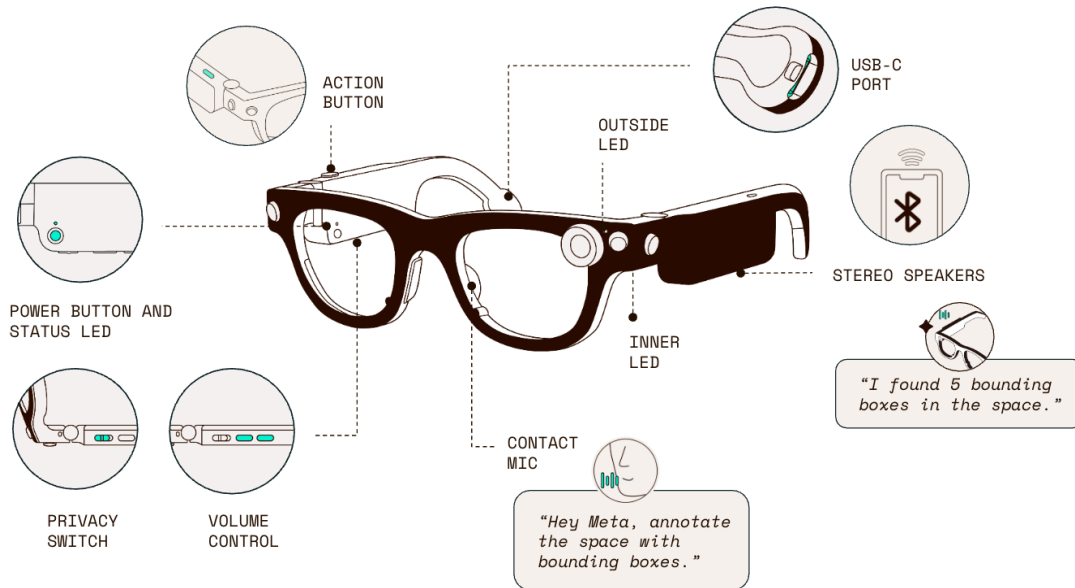


Figure 2b. Illustration of the interactivity features of the Aria Gen 2 device

Aria Gen2 introduces a state-of-the-art sensor suite that expands the range of research applications:

CV Cameras	Four computer vision (CV) cameras provide a wide field of view and high-dynamic-range (HDR) imaging, enabling robust scene understanding and perception even in challenging lighting conditions.
RGB Camera	RGB point-of-view (POV) camera for high-resolution egocentric vision and ML applications
Eye Tracking system	Enable precise gaze estimation and user intent modeling.
Spatial Microphones	Capture high-fidelity audio for spatial awareness and voice interaction.
IMU, Barometer, Magnetometer, GNSS	Support accurate localization, motion tracking, and environmental sensing.
Ambient Light Sensor (ALS) with UV Channel	Distinguishes between indoor and outdoor lighting, opening new avenues for context-aware research.
PPG Sensor (Nosepad)	Measures heart rate, supporting physiological and affective computing studies.
Multi-device time alignment	Using SubGHz radio based custom protocols allow for multiple Aria Gen 2 devices to align their timebases to 10's of microseconds accuracy.

*Table 2. Summary of sensors in Aria Gen 2 device.*

These new sensors, especially the ALS with UV channel, contact-mics and the physiological sensors in the nosepad, unlock research opportunities in context-aware computing, health monitoring, and robust multimodal perception.

## Hardware Specifications

Category	Capability
Sensors	4 CV cameras, RGB camera, 2 eye tracking cameras, 7 spatial microphones, 2 IMUs, barometer, magnetometer, GNSS, ALS (with UV), PPG, contact microphone.

Speakers	Two force cancelling speakers.
Coprocessor	Features Meta’s custom, energy-efficient coprocessor optimized for low-power, on-device AI workloads. <a href="#">Meta’s custom coprocessor blog</a>
Battery Life	Supports 6–8 hours of continuous recording use with a custom lithium-ion battery.
Multi-device time alignment	Custom radio protocol based on Sub-GHz radio technology for multi-device time alignment.
PPG Sensor (Nosepad)	Measures heart rate, supporting physiological and affective computing studies.
Connectivity	WiFi6E and BT 5.3 with 2 antenna chains.

*Table 3. Specification of the hardware attributes of the Aria Gen 2 device.*

### RGB camera:

The Aria Gen 2 features a high-resolution RGB camera with rolling shutter, integrating the Sony IMX681 sensor. This module provides a 133° horizontal and 99° vertical field of view, with a 12 megapixel, 1.0 μm pixel sensor capable of full-resolution (4032 × 3024) capture at 24 fps. The RGB pipeline supports functions such as binning, downscaling, and cropping, enabling higher frame rates and diverse video formats and output resolutions. A dedicated on-device color ISP is included within the RGB pipeline, offering features like auto-exposure and auto-white balance.

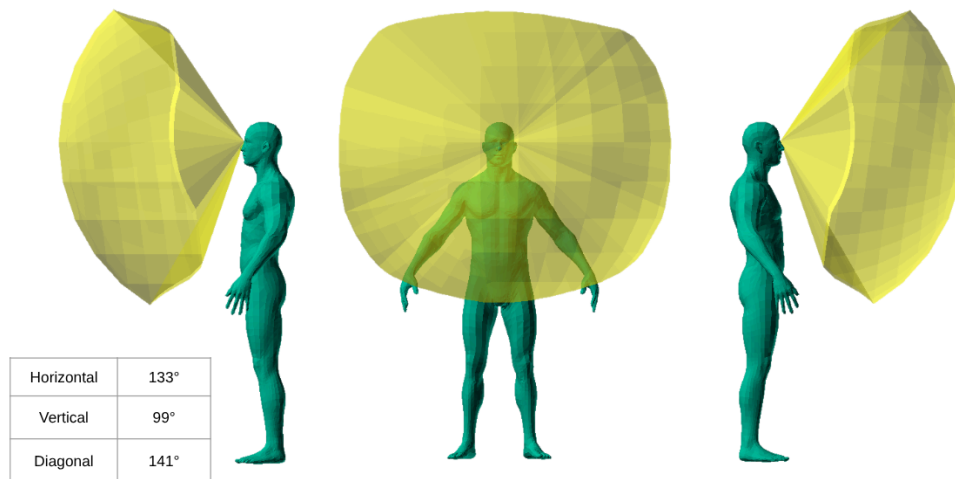


Figure 3. Field Of View (FOV) of the CV cameras on Aria Gen 2 device.

### Computer vision (CV) cameras:

Each of the CV cameras provides a field-of-view of  $119^\circ \times 119^\circ$  and contains a custom-designed [ref1, ref2], stacked digital pixel sensor with a global shutter. This sensor features  $4.6 \mu\text{m}$  pixels in a  $512 \times 512$  array and employs an overlapped triple quantization scheme to enable single-frame, single-exposure high dynamic range (HDR) imaging, with dynamic range capability exceeding 110 dB. Monochrome image data from each sensor is processed by a dedicated image signal processor (ISP), which performs noise reduction—including fixed pattern and defective pixel correction—and implements a tuneable HDR tonemapping scheme to map the high dynamic range output to standard 8-bit formats. The front-facing CV cameras form a stereo pair with significant overlap that enables off-device depth reconstruction of the scene.

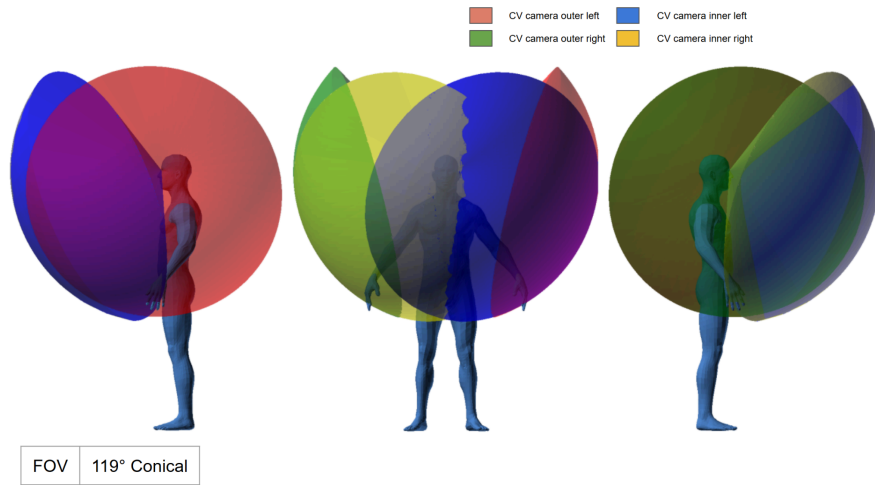


Figure 4. Field Of View (FOV) of the CV cameras on Aria Gen 2 device.

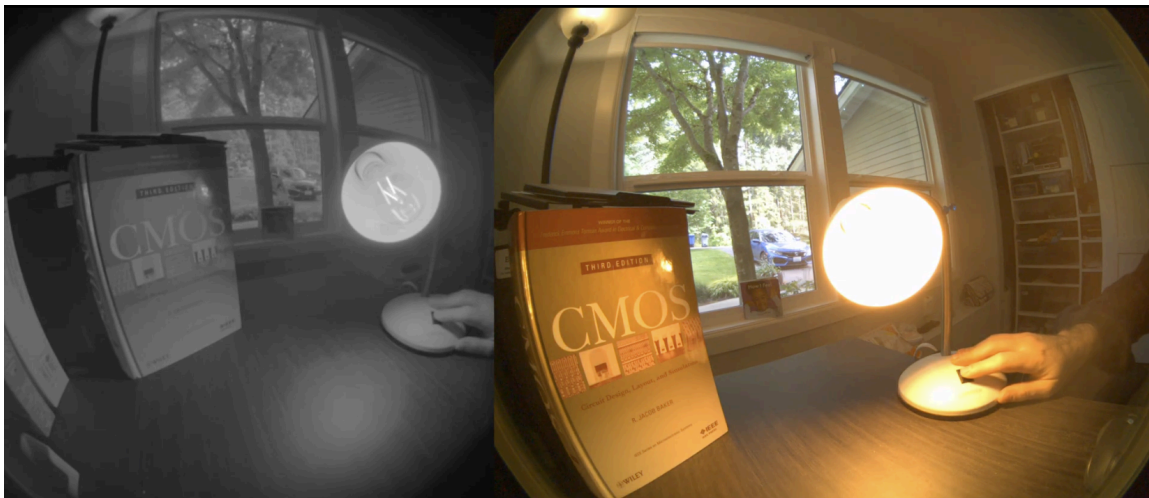


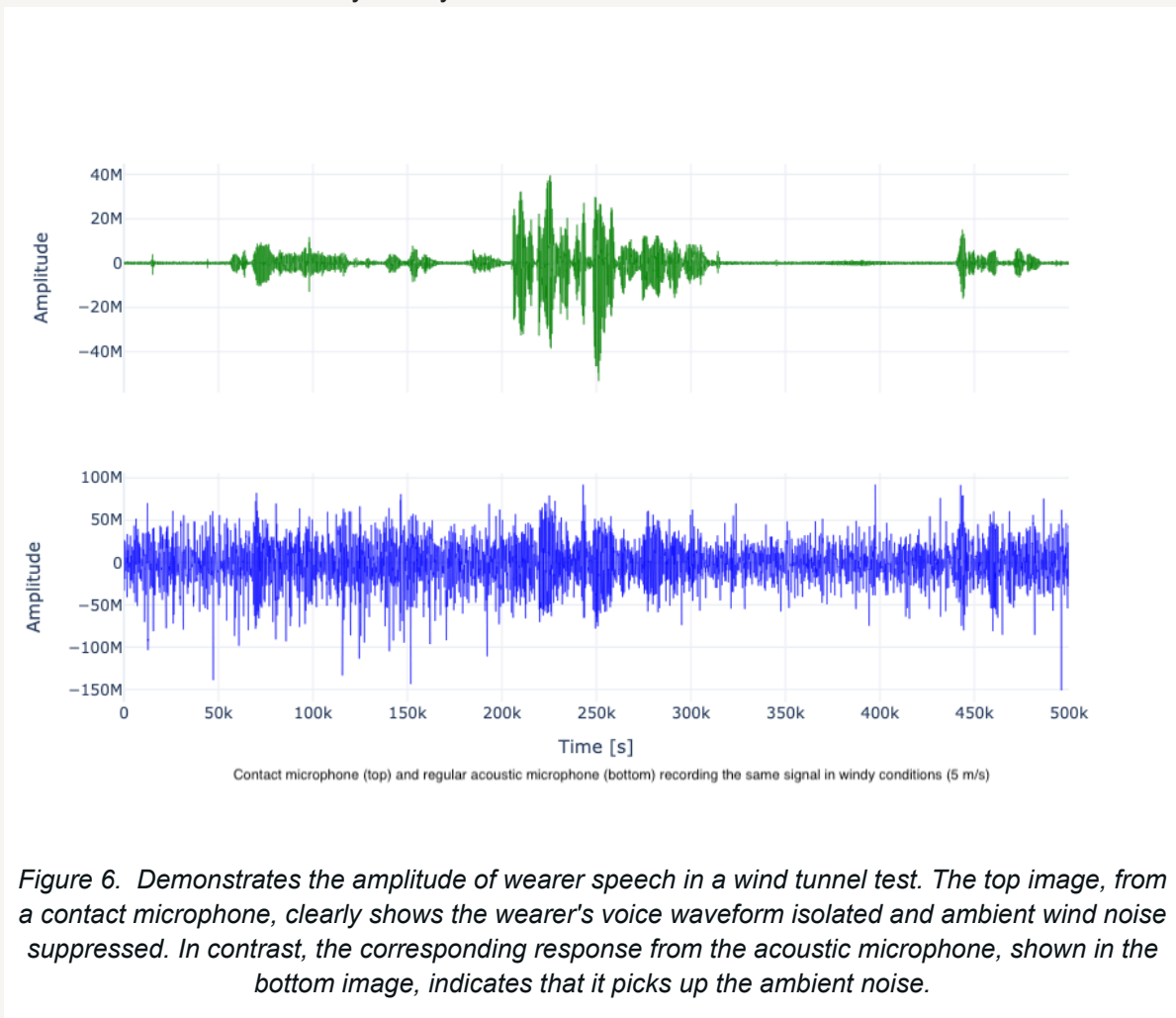
Figure 5. Comparison of the dynamic range of the CV camera (left) and RGB camera (right) that showcases the High Dynamic Range (HDR) capability of the CV camera.

### Spatial microphones:

The audio capture capabilities support 7 acoustic microphones at sampling rate of up to 48 KHz for each individual microphone. This high sampling rate ensures detailed and accurate audio recording, capable of capturing a broad spectrum of sounds with exceptional fidelity.

### Contact microphone:

The Aria Gen 2 device incorporates a high-fidelity contact microphone, the Knowles V2S200D, embedded within the glasses' nosepad. This placement is critical for optimal audio capture, minimizing external noise interference and maximizing the clarity of the user's voice. The microphone samples audio at up to 48 KHz, enabling clear capture of the wearer's voice even in noisy, windy conditions.



### Eye tracking system:

The eye tracking system for each eye consists of a 400x400 pixel resolution infrared (IR) camera and 8 IR LEDs. The data from these cameras is processed on-device, yielding

real-time, high quality gaze estimations at up to 90 Hz framerate. Alternatively, the data from the eye tracking cameras can be recorded and processed off-device.

#### PPG:

A photoplethysmography (PPG) sensor is integrated into the nosepad of the glasses and enables recording the wearer's heart rate. The PPG sampling rate is typically 128 Hz, with a maximum of 512 Hz.

#### Inertial measurement units (IMUs):

Aria Gen 2 features dual 6-axis gyroscopes and accelerometers with a maximum sampling rate of 1600 Hz and a typical sampling rate of 800 Hz. The data is consumed by the on-device VIO system and can also be recorded.

#### Barometer:

The barometer on Aria Gen 2 is a low noise sensor for atmospheric pressure with a maximum sampling rate of 240 Hz max and a typical sampling rate of 100 Hz.

#### Magnetometer:

The magnetometer on Aria Gen 2 is an ultra-low noise sensor with a maximum sampling rate of 400 Hz and a typical sampling rate of 100 Hz.

#### Proximity Sensors:

A proximity sensor detects when the glasses are worn. This sensor uses a threshold-crossing interrupt mechanism to communicate with the integrated computing system.

#### GNSS:

Global Navigation Satellite System (GNSS) support across bands such as L1, L1 + L5, E1, and E1 + E5. The device supports L1 + L5 (GPS) and E1 + E5 (Galileo) dual-frequency signals, which significantly improve acquisition speeds and positioning precision.

#### Ambient Light Sensor (ALS)

The VD6281 sensor provides correlated color temperature (CCT) and lux measurements. It features 5 channels (red, green, blue, IR, UVA), with an additional clear channel for flicker detection. We provide an example of ALS data below, captured during a recording as the wearer moves through an office environment and onto an outdoor balcony. The data highlights three frames, demonstrating that the UV and IR channels of the ALS sensor exhibit distinct changes when transitioning to outdoor environments. These signals can be leveraged for classification of indoor/outdoor environments.

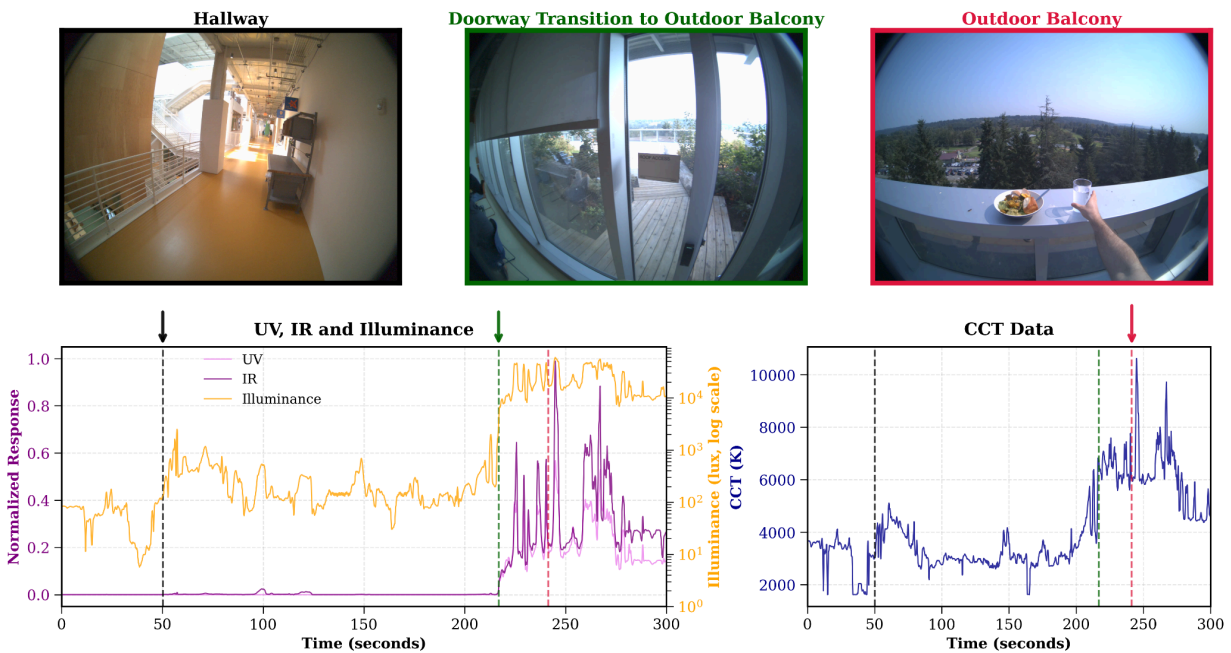


Figure 7. Illustrates the measurement from the various channels of the ALS across different scenes from indoor to outdoor scenarios. The data highlights three frames, demonstrating that the UV and IR channels of the ALS sensor exhibit distinct changes when transitioning to outdoor environments and can be used for classification of indoor/outdoor environments.

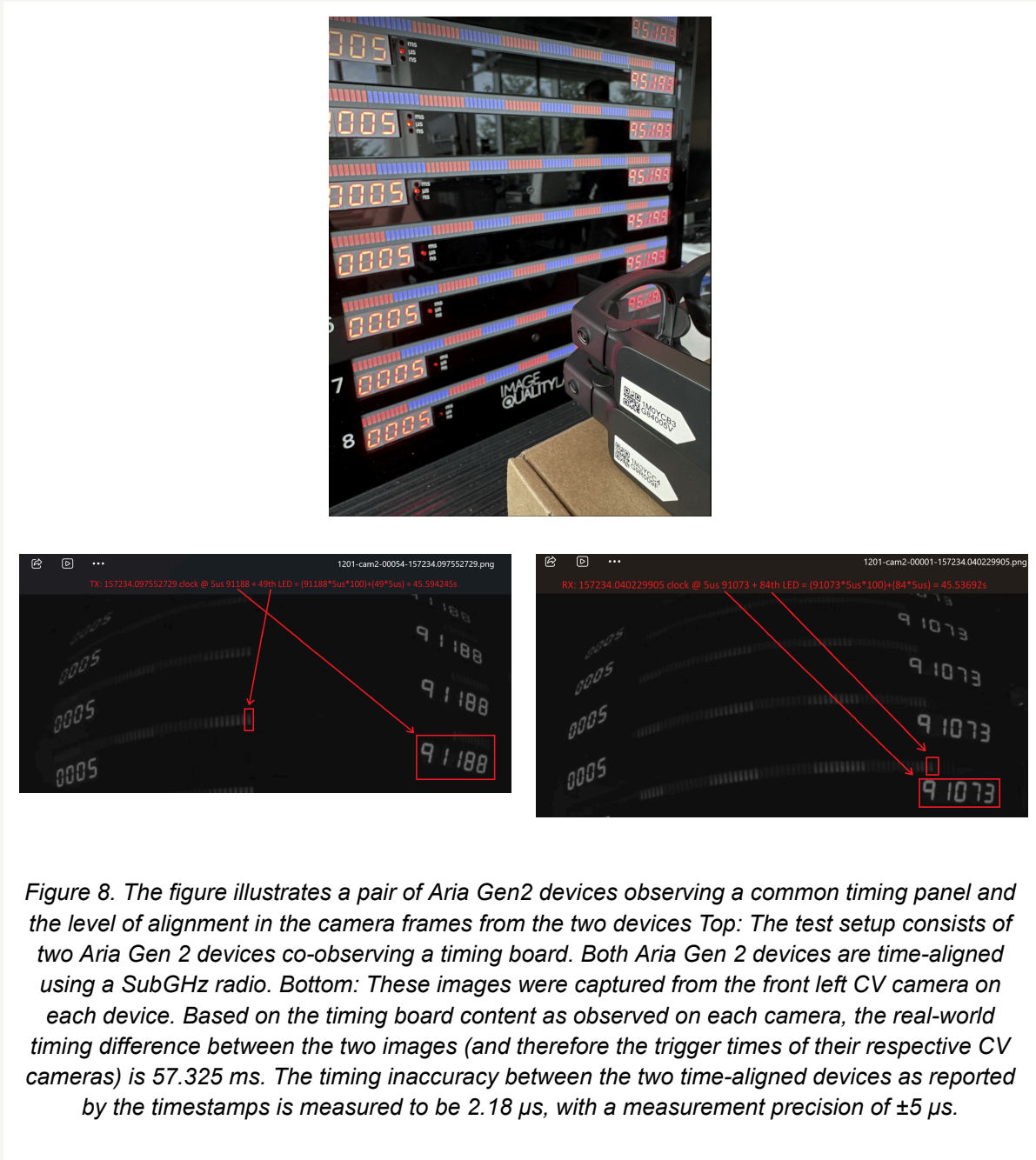
## On-device hardware accelerators

The device's coprocessor supports on-device compression for image/video using H265 HEVC and audio data using OPUS encoders. It also features hardware-accelerated support for machine perception, including 3D articulated hand tracking, eye-tracking with gaze per eye output, and advanced signals such as pupil diameter and blink detection, as well as 6DoF localization.

## Multi-device time alignment

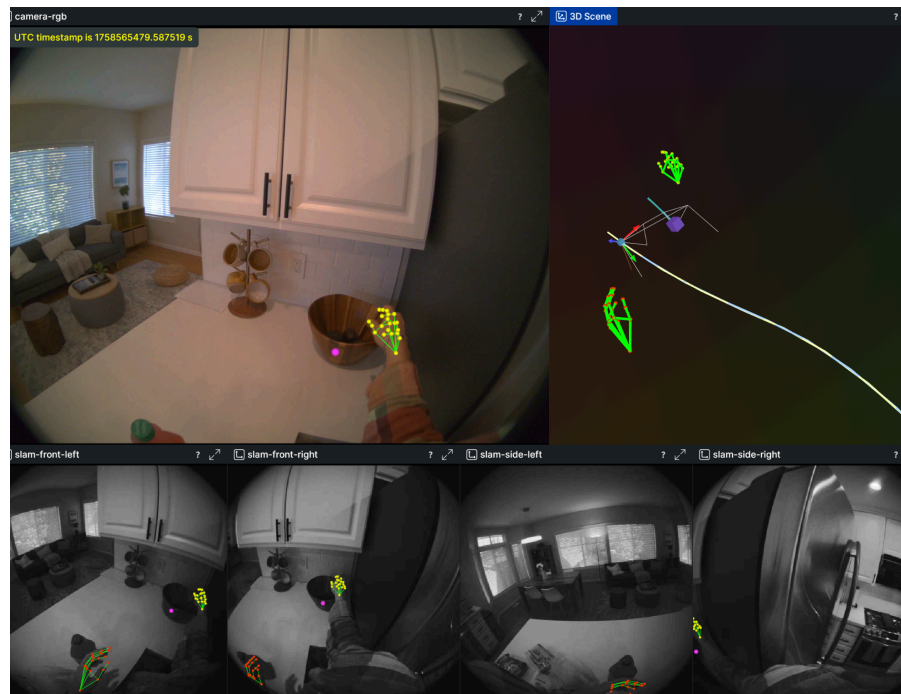
A SubGHz radio facilitates the transmission of device timestamps between devices. One broadcaster device transmits its time, while other devices receive this broadcast and subsequently compute time alignment (difference) between the broadcaster and receiver. (Note that this does not synchronize the devices.) This approach circumvents the need for round-trip communication typically associated with TicSync. The system operates under the assumption of negligible time-of-flight within its 30m indoor and 100m outdoor operating ranges. The measured time offset error has been observed to be less than 10  $\mu$ s -- roughly

2 orders of magnitude smaller error than the reported accuracy of TicSync of ~1ms offset error [citation]. Methodologies for integrating Aria Gen2's time alignment capabilities with other systems may be made available in the future.



## On-device software capabilities

The Aria Gen2 device has energy efficient hardware acceleration in the coprocessor that enables on-device machine perception algorithms to be run with all-day wearability in mind. The output of the machine perception algorithms output can be either recorded onto the device or streamed to a host PC. Profiles are used to precisely configure the frequency and types of machine perception algorithm calculations for a variety of research applications. At the time of the release of Aria Gen 2, the device is able to run visual-inertial odometry (VIO), 21-keypoint hand tracking, eye gaze estimation, and automatic speech recognition algorithms. However, there is significant potential to expand what can be computed on-device in the future, such as algorithms for scene understanding, action recognition, human body pose estimation, and more. We will now describe the currently available machine perception algorithms in detail.



*Figure 9. This visualization displays the results of all on-device machine perception algorithms streamed from the device, alongside the RGB and CV cameras. The top-left and bottom figures illustrate the RGB and CV cameras, respectively, with overlaid eye-gaze (pink dot) and articulated joint positions (yellow dots). The top-right figure presents a 3D scene depicting eye-gaze (blue line segment), hand-poses (green hand skeleton), and VIO trajectory (white/blue line). All of these signals are computed in real-time on the device.*

## Visual Inertial Odometry (VIO)

One of the key features of Aria Gen 2 is its ability to track the glasses in six degrees of freedom (6DOF) within a spatial frame of reference using Visual Inertial Odometry (VIO), by fusing the sensor data from the four CV cameras and the two IMUs. This allows for seamless navigation and mapping of the environment, opening up new possibilities for research in contextual AI and robotics.

The VIO output is generated at 10Hz with the following output:

- 3-DOF position
- 3-DOF linear velocity
- 3-DOF orientation in quaternion form
- 3-DOF angular velocity
- Estimated direction of gravity for the odometry frame

Additionally, Aria Gen2 also produces high-frequency VIO output (the fields of the output are the same regular VIO) at IMU rate (800Hz), by performing IMU pre-integration on top of the regular 10Hz VIO output. The high-frequency VIO output can be useful for applications where low-latency VIO poses are needed.

## Eye Tracking

Aria Gen 2 also contains an advanced camera-based eye tracking system that tracks the wearer's gaze. The advanced gaze signal enables a deeper understanding of the wearer's visual attention and intentions, unlocking new possibilities for human-computer interaction. This system generates the following eye tracking outputs for **each eye**, up to 90Hz:

- The origin and direction of the individual gaze ray
- The 3-DOF position of the entrance pupil
- The diameter of the pupil
- Whether the eye is blinking

Additionally, the system also produces the following signals for the **combined gaze** estimated from both eyes, including:

- The origin and direction of the combined gaze ray
- Vergence depth of the combined gaze
- Distance between the left/right eye pupils, a.k.a, IPD

## Hand Tracking

Aria Gen 2 also features a hand detection and tracking solution that tracks the wearer's hand in 3D space. This produces articulated hand-joint poses in the device frame of

reference, facilitating accurate hand annotations for datasets and enabling applications such as dexterous robot hand manipulation that require high precision.

The hand tracking pipeline generates the following outputs at 30Hz for each hand (left and right):

- 3-DOF position of the wrist
- 3-DOF rotation of the wrist
- 3-DOF positions of the 21 finger joint landmarks

## Automatic Speech Recognition (ASR)

Aria Gen 2 is capable of running on-device ASR real-time for real-time speech transcription of the wearer. This allows interactive applications such as conversational AI chatbot to be built where users' speech commands can be parsed and processed. The spatial acoustic microphone array on the glasses are leveraged to run beamforming and side-speaker speech suppression, and the beamformed audio stream is fed to the ASR pipeline to transcribe only the wear's voice. We also run an endpointing model to identify when a speaker stops talking.

## III SDK Overview

Similar to the Aria Gen1 SDK, the Gen 2 Software Development Kit (SDK) is a comprehensive suite of tools designed to facilitate robust configuration and control of Aria Gen 2 glasses. The SDK provides both Python libraries and command-line utilities, enabling seamless device management, secure authentication, and efficient control of recording and streaming functionalities. Through a unified interface, researchers can initiate and manage high-fidelity time-synchronized multi-modal sensor data capture—including camera, audio, inertial data, and machine perception streams such as visual-inertial odometry (VIO), hand pose, and eye gaze—while ensuring data integrity and security. It also provides access to intrinsic and extrinsic sensor calibrations, device metadata, and network synchronization utilities that ensure accurate data fusion for downstream research

The SDK emphasizes ease of data-collection and streaming for research workflows: It supports automated scripting for large-scale data collection, as well as real-time visualization and monitoring of both recorded and live-streamed data. Example code and detailed documentation further accelerate onboarding and experimentation. Please refer to the [SDK](#) and its [technical specification](#) for more details.

## IV Use Cases and Applications

A large portion of research published with Aria Gen 1 focused on the recording of egocentric data for the training and benchmarking of AI models and spatial algorithms. Notable example datasets include [HOT3D](#), [Nymeria](#), [Ego-Exo4D](#), [HD-EPIC](#), [EgoLife](#), and [LaMARIA](#). Others have used Aria’s real-time streaming capabilities to run live prototypes that receive and process the device’s sensor data as input into robot control policies as seen in [EgoMimic](#) and egocentric assistants as seen in “[Egocentric expert guidance for procedural activities using Smart Glasses, Computer Vision and RAG](#)”.

Aria Gen 2 will transform the way egocentric research is performed. Additional sensors in Aria Gen 2, including a PPG sensor, additional computer vision cameras, contact microphone, and ambient light sensor will open up the possibility for the development of richer multimodal egocentric datasets that capture more of the wearer’s experience, state, and environment. On-device machine perception will make the development of datasets with Aria Gen 2 more efficient, often avoiding the need to use our offline Machine Perception Services (MPS). It will also make real-time streaming applications more feasible, which previously required processing of streamed data on a separate machine for real-time results. Enabling interaction and closing the loop with the user via audio will enable expanded research in the area of human-computer interaction, human-robot collaboration, and assistive technologies such as those demonstrated in partnership with Envision in the [spatial-audio navigation case study](#). Interaction capabilities may also open up new kinds of dataset development, where the wearer can interoperate with models and live systems during recording.

In the field of Contextual AI, we anticipate that as LLMs and VLMs grow in capability, researchers will increasingly rely on real-time testing of their functionality with Aria Gen 2. We also foresee that Contextual AI datasets will become longer in duration and more naturalistic as wearers become accustomed to long-duration recordings throughout their normal days. This will unlock the development of AI agents capable of long-range reasoning with highly detailed personal context. In robotics, we will see Aria Gen 2 used as a human data collection tool for imitation learning, a real-time human action capture tool for teleoperation, and as a vehicle for robotics systems to understand a human wearer’s state and actions for human-robot collaboration development.

## V Conclusion

Aria Gen 2 represents a significant evolution in wearable multimodal sensing for egocentric research and contextual AI. Building on the success of Aria Gen 1, the new platform advances the boundaries of real-world data capture and real-time machine perception. Through an expanded sensor suite, improved ergonomics, and energy-efficient on-device computation, Aria Gen 2 enables long-duration, high-fidelity recording and streaming of

multimodal perception signals—including video, audio, inertial, gaze, and physiological data—within a lightweight, all-day wearable form factor. These technical improvements dramatically reduce the friction of dataset collection and open up new classes of research that integrate perception, embodiment, and user context.

By combining advanced hardware with a robust SDK, Aria Gen 2 provides a reproducible and scalable framework for researchers to study real-world AI systems in situ. The SDK's unified APIs for device control, synchronization, and data management streamline the entire research workflow—from experimental setup to real-time visualization and large-scale dataset generation—bridging the gap between algorithmic development and embodied deployment. Together, these capabilities make Aria Gen 2 not only a successor to its predecessor, but a foundational platform for exploring long-horizon, interactive, and context-aware AI systems.

Looking ahead, we envision Aria Gen 2 catalyzing a new generation of research in egocentric perception, human-computer interaction, and robotics. Its ability to sense, process, and respond to the world in real time provides a tangible step toward embodied intelligence—AI systems that understand humans and their environments through shared context and experience. As the research community builds upon this platform, Aria Gen 2 will help accelerate progress toward truly personal, adaptive, and contextually-aware augmented-reality computing.