L+R

Strategy+Aesthetics

# Ephemeral Perception

## A Developer's Framework for Building Privacy-by-Architecture Ambient AI and Smart Glasses Apps

A reference architecture for developers building on Meta's Device Access Toolkit, Google's Android XR, and the broader ambient computing ecosystem.

CREATED BY

ALEX LEVIN + IVAN LEIDER

WHITE PAPER VERSION 1.0

JUNE 2026

# Authors

**Alex Levin** is the co-founder of L+R, a SOC 2 Type 2 certified international digital product studio working at the intersection of strategy, design, and emerging technology for Fortune 500 companies, small to mid-sized businesses, and ambitious startups. Levin has been an emerging voice in mobile applications' role in ambient computing and spatial technology, presenting to bipartisan United States Senate and White House staffers on the real-world implications of AI and wearables, and speaking at conferences including CES, Mobile World Congress, and Web Summit. He has written on technology and design for Forbes and has been featured in the Wall Street Journal. At Meta Connect 2025, L+R was named alongside Disney, Logitech, and Microsoft as a launch partner for Meta's Device Access Toolkit, making L+R one of the first third-party studios globally with developer access to the camera and audio capabilities of Meta smart glasses. L+R has since extended this relationship into the next generation of Meta's wearable platform, partnering with Meta to bring developer-built web applications to Meta Ray-Ban Display, Meta's first consumer smart glasses with an integrated visual interface.

**Ivan Leider** is the Director of Engineering at L+R, where he has led the spatial computing, web, mobile, and emerging technology practices for over nine years. His career in mobile development stretches back to the launch of the App Store in 2008 and includes foundational work on early Bluetooth-connected iOS hardware integrations. He has overseen the development of cross-functional platforms used by millions of people worldwide. Leider was selected to represent L+R at Apple's invitation-only Vision Pro Developer Labs in London and was invited by Meta into the alpha and beta phases of the Device Access Toolkit. His firsthand experience with the toolkit's capabilities and constraints directly shaped the technical grounding of this framework.

## Technical Contributors & Advisors

**David Levin** is a retired Lead Systems Engineer and cybersecurity pioneer whose five-decade career provided the foundational security philosophy for this framework. During his twenty-two years at BBN Technologies, he served as a Principal Investigator within the Cyber Security business unit, leading rigorous information assurance projects including the DARPA Brandeis mobile privacy initiative. His historical contributions to the computing industry range from early work on the Multics operating system at Honeywell in the 1970s, to developing the bytecode interpreter that enabled VisiCalc to launch alongside the original IBM and Digital PCs, to pioneering Kerberos-based enterprise single sign-on at Hewlett-Packard. His lifelong dedication to secure system architecture, digital forensics, and privacy directly inspired both the architectural principles of Ephemeral Perception and the trajectory of L+R.

**Ilan Levin** is a hardware engineer and technical program manager with twenty years of experience delivering complex sensor, robotics, and embedded systems. Currently a Senior Program Manager at Collaborative Robotics, his career spans defense-grade sensing systems at Raytheon BBN Technologies, hardware engineering at Newell Brands' XO Innovation Lab, and large-scale hardware deployment at Masabi. Levin contributed the hardware architecture validation and sensor system perspective that informs the Perception Layer of this framework.

## Acknowledgments

The author and contributors would like to thank the L+R engineering and design teams whose hands-on work building on these platforms generated the technical insights at the heart of this framework, and the broader mobile, spatial computing, and wearable technology developer community whose questions and conversations helped sharpen these ideas. Thanks also to the participants of Meta's Wearables Community Summit, Meta Connect, Apple WWDC, Google I/O, Mobile World Congress, 4YFN, Web Summit, FlutterCon, Droidcon, the Meridian Meetups community, the World AI Eyewear Alliance, and L+R's Emerging Technology Accelerator Program for the ongoing dialogue that informed this work.

## Disclaimer

L+R is a launch partner for Meta's Device Access Toolkit. This relationship is disclosed transparently and has informed, but does not compromise, the independence of the framework and analysis presented here. Nothing in this paper reflects confidential or proprietary information. All technical claims are based exclusively on publicly available information and L+R's own experience working with publicly accessible tools and SDKs. The views expressed reflect the author's own perspective and do not represent the official position of Meta or any other platform or hardware manufacturer referenced in this document. This framework is published as a reference architecture and is not proprietary to L+R.

# Table of Contents

# Executive Summary

We are at an inflection point in computing. For decades, the primary interface between humans and digital systems has been the screen, a rectangle we look down at, carry in our pockets, and surrender our attention to. That era is ending. A new generation of wearable devices including AI smart glasses, ambient AI pendants, smartwatches, connected rings, wristbands, and more, is placing computation directly in the field of human vision. The screen is disappearing, and the world itself is becoming the interface.

This transition carries extraordinary promise, and it also carries a structural risk that the industry has not yet named clearly enough to address.

The dominant architecture for ambient AI wearables today is built on a default we call Capture-Store. Devices collect raw video footage and audio, transmit it to cloud infrastructure for processing, extract meaning from it, and return a response to the user. This approach works functionally, but it creates an unavoidable chain of custody for raw data. It captures and stores the environments, faces, conversations, and private moments of every person in the wearer's field of view, whether they consented or not. It treats raw visual data as a resource when, in ambient computing, it is a liability.

To date, the Capture-Store architecture was tolerated as an implicit trade: for example, users surrendered their intentional web searches for free services, and companies profited by hoarding and monetizing that data. Today, however, continuously and arbitrarily capturing the physical world breaks that historical contract, crossing a dangerous line into violating personal privacy, which becomes an immediate regulatory liability.

To be clear, the troubling effects of this architecture are largely unintentional. They are simply the result of applying legacy smartphone and cloud architectures to a completely new medium. But defaults, once established at scale, become standards. By the time regulators arrive, the habits of an entire developer ecosystem will already be formed.

This white paper proposes an alternative, more powerful solution, without the endless and unnecessary collection of potentially harmful data.

We call it **Ephemeral Perception**.

Ephemeral Perception is a design philosophy and reference architecture for both ambient and session-based AI hardware and software. It dictates that raw visual data be converted into structured semantic information at the point of capture and discarded immediately and irreversibly. The device perceives the world, the meaning is extracted, and the raw visual signal

is destroyed. What remains is useful, structured, and private by architecture rather than by policy.

The Ephemeral Perception framework did not emerge from abstract theory or policy debates. It emerged from the trenches of product development: writing user stories, defining PRDs, designing UX/UI, writing raw code, prototyping real-world use cases, and navigating the actual friction of building for these devices.

Far from trading capability for safety, this hands-on experience proves the opposite. Because Ephemeral Perception runs inference on-device, it eliminates the cloud round trip that adds latency and the continuous radio streaming that drains the battery; and because it emits a lightweight semantic stream rather than a single exclusive video feed, that output can be consumed by multiple applications at once. The result is lower latency, lower power draw, and broader utility, achieved by architecture rather than by trade-off. Privacy becomes a natural result of the architecture itself, rather than a point of compromise or an added layer of policy enforcement.

The conversations that shaped this framework come from stakeholders across the growing ambient computing ecosystem: developers encountering regulatory resistance to Capture-Store systems, eyewear and hardware companies racing to embed smart technology into wearables, and platform owners asking the same questions about use cases, hardware criteria, and how to build applications users will actually trust wearing all day. Ephemeral Perception is the direct architectural answer to those pressures.

This white paper makes the case for a more powerful, user-friendly architecture for edge and on-device processing and the desire for its widespread use, before Capture-Store calcifies into the industry default. This framework provides the credible privacy, accuracy, performance, and battery story that hardware companies and developers desperately need. To be clear, cloud processing still plays an important role for these devices, and for many use cases, it remains the right choice. But it should never be the initial processing step, nor should it be the go-to solution.

# I. Introduction: The World Beyond the Rectangle

The history of personal computing is a history of the screen. From the first monitors to the smartphone in your pocket, the dominant paradigm has required you to redirect your attention away from the world in front of you and toward a luminous rectangle. The information lives there, and you go to it.

Smart glasses and ambient AI wearables represent the first serious architectural challenge to that paradigm in the history of consumer technology. Rather than requiring you to look down, they allow you to look forward, with computation coming to you, layered directly onto the physical world you already inhabit.

But this shift introduces a profound vulnerability. To layer information onto the physical world, these devices must constantly observe it. The camera is no longer just a lens; it is a continuous sensor of the user's reality. As ambient wearables scale from niche gadgets to everyday utilities, the default hardware architecture of the last decade has become fundamentally incompatible with human privacy.

This vulnerability is compounding rapidly. Academic and industry researchers are already prototyping the next evolution of this technology: always-on wearable agents that integrate continuous visual perception directly with autonomous task execution.[1] [2] These systems are designed to constantly scan a user's environment, read physical documents on their desk, and automatically trigger web actions, such as creating calendar events or adding items to a shopping cart, without manual input. If these agentic systems are built on the legacy Capture-Store architecture, the privacy implications are catastrophic. It means raw, continuous video of private spaces is not just being stored; it is being actively streamed into autonomous pipelines.

We cannot solve this with faster pipelines, bigger cloud servers, better consent forms, stricter cloud security, or updated privacy policies. The vulnerability is the raw data itself. Every raw frame that exists beyond the moment of capture can be considered a liability: a privacy risk, a storage cost, a regulatory exposure, and a performance bottleneck. To secure the ambient computing era is to re-architect how devices see the world, extracting the semantic asset and destroying the visual liability.

This is a disruptive shift, not an incremental one, and like all genuine shifts in computing, it is arriving faster than the frameworks needed to govern it responsibly. Consider what has happened in just the past few months:

- **Meta sold over 7 million smart glasses in 2025 alone**, capturing the dominant share of the global smart glasses market. Following this massive hardware adoption, their Device Access Toolkit, the first SDK to give third-party developers direct access to the camera and audio of those millions of devices, entered active developer preview following Meta Connect 2025.[3] Meta has since opened Meta Ray-Ban Display, its first consumer smart glasses with an integrated visual interface, to third-party web applications, dramatically expanding what developers can build on the platform.

- **Apple's impending entry** became an open secret in early 2026, with credible supply-chain reports and software leaks pointing to a late 2026 preview of their first premium smart glasses, heavily integrated with their new "Visual Intelligence" framework.

- **Google confirmed Android XR**, its open platform for smart glasses, and used Google I/O 2026 to detail the full developer roadmap, including expanded SDK access, deeper Gemini integration for on-device perception, and confirmed hardware partnerships with Samsung, Warby Parker, Gentle Monster, and Kering Eyewear. Samsung began quietly laying the software groundwork for its upcoming 'Galaxy Glasses' in its latest developer betas.[4]

- **Qualcomm unveiled the Snapdragon Wear Elite**, the first wearable chip with a dedicated neural processing unit capable of running two-billion-parameter AI models entirely on-device.

- **CES and MWC 2026 saw an explosion of over 250 exhibitors** in the XR space, highlighted by companies unveiling eSIM-enabled smart glasses, proving these devices are rapidly evolving from tethered accessories into standalone computers.[5] [6]

- **Alibaba** announced global availability for its Qwen AI Glasses, further accelerating the global hardware race.[7]

The market is projected to reach 20 million units and $5.6 billion in value by the end of 2026.[8] Like all disruptive shifts, the ambient computing era has arrived chaotically. It arrived all at once, and with a hodgepodge of architectures, frameworks, and approaches; not with a carefully designed and shared framework for how the developers, hardware manufacturers, and platform owners building this ecosystem should handle its most sensitive byproduct: the continuous visual stream of the physical world.

# II. Capture-Store and Where It Breaks Down

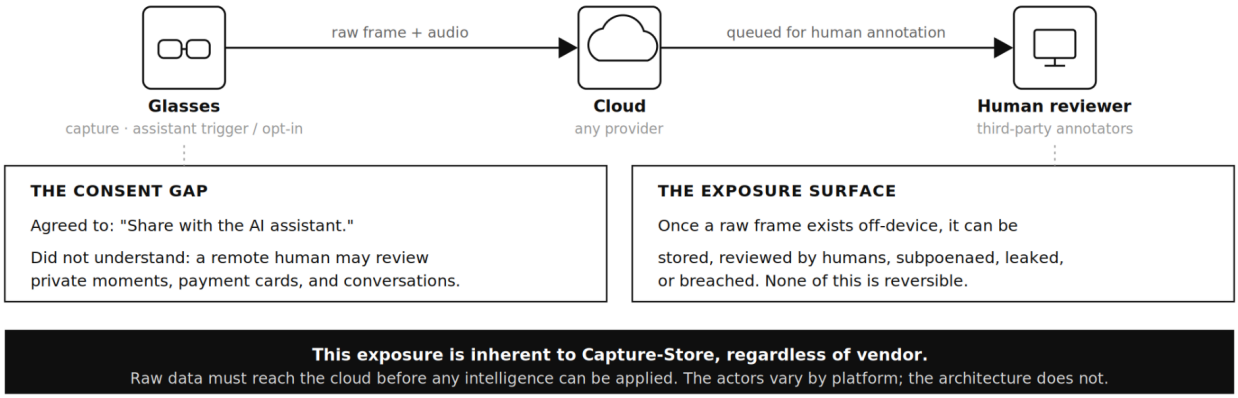
Every major platform currently operating in the ambient AI space inherited the Capture-Store architecture from an earlier era of connected devices. Back then, on-device processing was too limited to do the heavy lifting, making cloud infrastructure the only viable option for running sophisticated AI models. The chain is simple: capture raw visual and audio data, transmit it to the cloud, process it there, and return a result. Beyond these technical bottlenecks, Capture-Store produces a severe and now well-documented ethical failure mode.

When our team at L+R began building on Meta's Device Access Toolkit, and later on Meta Ray-Ban Display, the first challenge we encountered had nothing to do with privacy. It had to do with how this prevailing architecture was fundamentally unsuited to the device. The architecture assumed that raw media capture was the primary goal, optimizing the hardware for continuous transmission rather than on-device perception. The camera feed was degraded, making text in the physical environment difficult to resolve accurately, and the battery drained faster than expected. The system was working exactly as designed, and that was precisely the problem. The device was fundamentally tuned for a capture-and-transmit pipeline. Because the system prioritized encoding and uploading heavy raw video files to the cloud, the local computer vision models were starved of the compute cycles necessary to perform accurate real-time character recognition.

This came to light in early 2026. An investigation by Swedish newspapers *Svenska Dagbladet* and *Göteborgs-Posten* revealed that human data annotators employed by Sama, a Kenya-based subcontractor for Meta, had been reviewing footage captured by Ray-Ban smart glasses.[9] [10] This included footage of people in private situations, credit card numbers, and personal conversations. The footage reached those reviewers through a specific chain: the wearer had opted into data sharing with Meta AI during initial setup, and then either actively recorded or invoked the AI assistant by saying "Hey Meta" to describe what they were seeing. The users had technically consented, but the reality of what that consent meant in practice (that intimate moments could end up in a remote human reviewer's queue) was not meaningfully understood by most.

The ensuing coverage focused heavily on Meta's corporate policies. However, the more important story is about the architecture itself. When a raw visual feed has to travel to a cloud server before any intelligence can be applied to it, that footage exists as a transmissible asset (Figure 1). It can be stored, reviewed, subpoenaed, leaked, or breached. This privacy risk is not a bug that can be patched with better consent language. It is a structural consequence of the design.

**FIGURE 1: The Capture-Store Chain of Custody**
Where raw visual data is exposed once it leaves the device




Meta's own recent advancements point directly toward the solution. In late 2025, Meta released SAM 3 (Segment Anything Model 3), one of the most capable computer vision systems ever built. Using a simple text prompt, SAM 3 can identify, segment, and track every object, person, surface, and concept across an entire video stream with remarkable precision and speed.[11] It does exactly what ambient AI wearables need to do: it converts raw visual input into structured, actionable semantic data. A street scene becomes a labeled map of relationships. A workspace becomes a parsed inventory of tools and people. A retail environment becomes a structured catalog of products and spatial positions.

SAM 3 proves our core thesis: the raw footage was never the actual goal. The structured meaning derived from that footage is what matters, and the technology to extract that meaning now exists at extraordinary levels of capability. The remaining gap is purely architectural. Because SAM 3 still operates at hundreds of millions of parameters, it currently runs on Meta's server infrastructure, meaning the raw footage still has to travel to the cloud before this distillation happens. The capability to make raw footage obsolete already exists. Now, the on-device architecture to execute it locally is finally arriving, and the developer community needs a framework for how to use it.

# III. Ephemeral Perception: The Framework

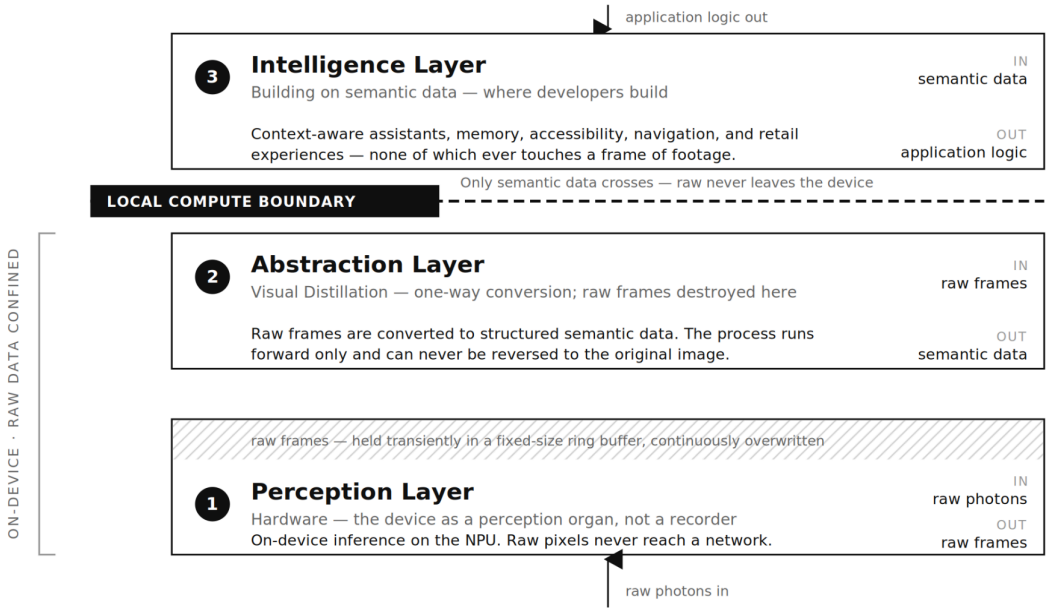
Ephemeral Perception is built on a single foundational claim: extract the meaning, without exposing the raw data. A computer vision system worn on a human face should operate the way human vision actually works. You perceive the world, you understand it (extract meaning). At no point is that dependent on generating a permanent visual record of it.

Ephemeral Perception introduces Consent-by-Architecture as its defining principle. Under this model, privacy relies on hard engineering constraints rather than human behavior. The framework does not require a user to decipher terms of service, a developer to write a flawless privacy policy, or a regulator to police data streams after the fact. Because the system's fundamental operation neutralizes raw visual data at the exact point of capture, data misuse becomes technically impossible at the code level. Privacy is achieved through structural design rather than legal enforcement.

The framework has three layers, each corresponding to a different level of the stack (Figure 2).

---

**FIGURE 2: The Three-Layer Framework**
How raw perception becomes structured meaning, and where raw data is confined




---

## The Perception Layer: Hardware

At the hardware level, Ephemeral Perception requires that visual inference happens on the device, so raw pixels never leave the device. The device is a perception organ, not a recorder. The volume of raw visual data a wearable camera generates is continuous and enormous. Managing it requires a preprocessing gate within the Perception Layer, one that operates like a lossy compression algorithm, holding resolution and sample rate as low as possible and scaling up only when something new enters the frame. In a static environment, the camera effectively sleeps. This triage function is what makes on-device inference practical: the system extracts what matters and discards its inputs so that no downstream layer is involved.
The hardware to support this is no longer theoretical. Qualcomm's Snapdragon Wear Elite, announced at MWC 2026, includes a dedicated Hexagon Neural Processing Unit capable of running AI models with up to two billion parameters entirely on device, without transmitting data to the cloud. Samsung, Google, and Motorola have already committed to building on this chipset.[12] The silicon for on-device visual distillation now exists and is shipping.

We call the design principle behind this layer Perceptual Sovereignty: a hardware architecture that keeps raw visual data in a fixed-size ring buffer, continuously overwritten as new frames arrive. This reduces on-board storage requirements and prevents its transmission outside of the local compute boundary. Raw frames exist transiently in this buffer for the duration of inference and are overwritten before any transmission boundary is crossed; the architectural guarantee is not that pixels never exist, but that they never leave. For enterprise deployments in particular, Perceptual Sovereignty addresses a significant and underappreciated concern. An employee wearing a camera-equipped device in a sensitive workplace environment, without a clear architectural guarantee that footage never leaves the device, creates meaningful exposure around trade secrets, client confidentiality, and regulatory compliance. Perceptual Sovereignty removes that exposure by design.

Perceptual Sovereignty applies to continuous ambient inference, the ongoing visual processing that makes a wearable AI assistant useful. It is not in conflict with intentional capture. When a user explicitly chooses to take a photo or record a video, that is a deliberate act with its own consent model, governed by the device's camera controls and the user's explicit intent. The architectural distinction matters: ambient inference operates under Ephemeral Perception principles by default; intentional capture operates under the user's explicit instruction. A well-designed Ephemeral Perception implementation makes this boundary clear and enforced. The system knows the difference between perceiving and recording, and behaves accordingly.

## The Abstraction Layer: Visual Distillation

The second layer is the philosophical and technical heart of the framework. We call it Visual Distillation: the irreversible, one-way conversion of raw visual input into structured semantic data at the point of capture.

Visual Distillation only works in one direction. It converts an image into structured meaning, but it cannot reconstruct the original image from its output. Visual Distillation creates an information stream, not visual recordings.

In practice, this means:

- Text in the frame becomes a sanitized string of characters. *Patterns like passwords, phone numbers, and credit cards are identified and dropped locally*
- A face becomes a presence signal, not an identity record
- An object becomes a semantic label and a spatial coordinate
- An environment becomes a structured map of relationships
- A gesture becomes an action event
- A product becomes a metadata tag

**User intent as an override.** Distillation is the default state, but user intent can override it. A query like 'Who is that person?' may allow raw pixels to leave the device when the on-device model cannot answer locally. This is not a contradiction of the framework; it is the framework operating as designed. Ambient inference defaults to ephemeral distillation. Explicit user-initiated queries operate under the user's own consent. The boundary between perception and intentional capture is enforced architecturally, not left to user vigilance.

The output of Visual Distillation is useful for every application a developer would want to build on an ambient AI platform. It's also, by construction, private. There's no footage to subpoena, no frames for a data annotator to review, no visual record of the intimate moments a person moves through while wearing the device.

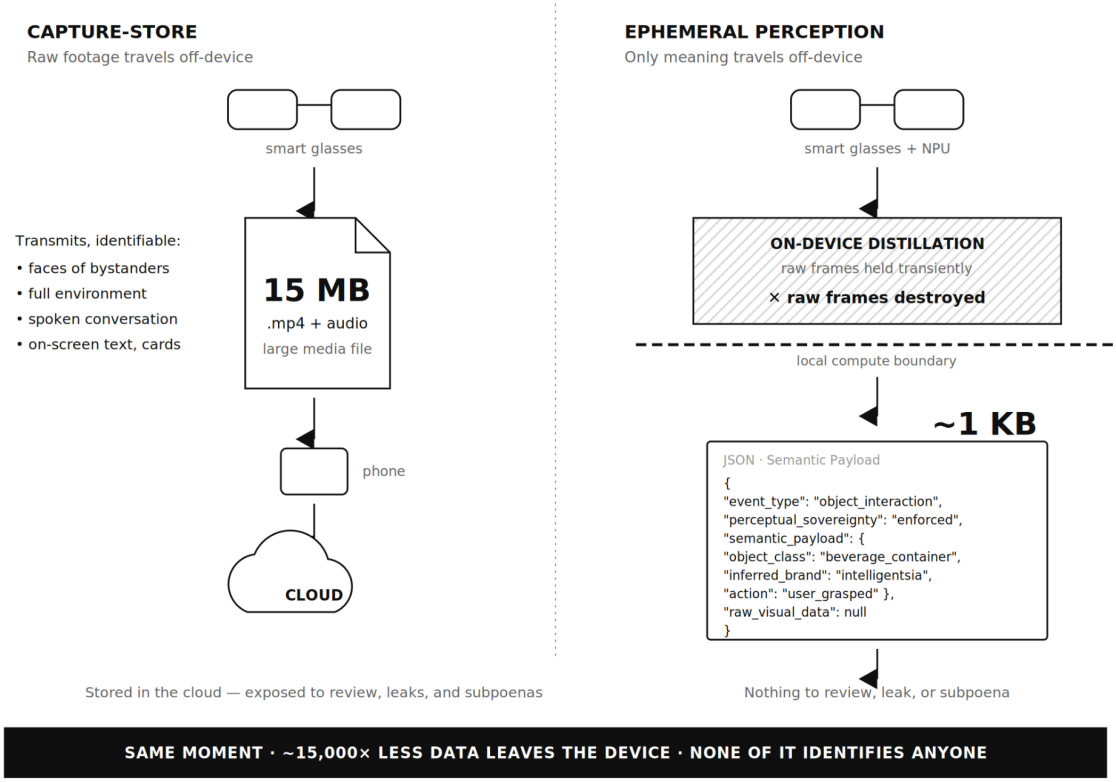
What remains after this process is what we call Semantic Payload.

To understand Semantic Payload practically, consider a user wearing smart glasses who picks up a bag of coffee. Under the Capture-Store architecture, the device transmits a 15MB MP4 video file to the cloud. Under Ephemeral Perception, the device distills the event, permanently destroys the raw footage at the compute boundary, and transmits only the Semantic Payload. This takes the form of an encrypted, lightweight JSON payload (Figure 3) that looks like this:

JSON


```
{
  "timestamp": "2026-04-12T14:32:01Z",
  "event_type": "object_interaction",
  "perceptual_sovereignty_status": "enforced",
  "semantic_payload": {
    "object_class": "beverage_container",
    "attribute": "coffee_bag",
    "inferred_brand": "intelligentsia",
    "action": "user_grasped",
    "relative_coordinates": [12.4, 4.2, 0.9]
  },
  "raw_visual_data": null
}
```

**FIGURE 3: What Leaves the Device**
The same moment under each architecture




This payload is directly usable by an application layer in ways a raw image file is not, since it is already structured, queryable, and bounded, and it carries a fraction of the identifiable risk. The Semantic Payload is not automatically free of Personally Identifiable Information (PII). Text strings, brand inferences, or location coordinates can still be sensitive. However, the surface area of this data is strictly bounded, inspectable, and governable in ways that raw video frames are not.

This deliberate limitation of risk is Consent-by-Architecture in practice. Because the structural design of the system guarantees that raw media never persists, developers and compliance teams can manage the remaining structured data using standard, highly auditable security protocols. The architectural guarantee established earlier is fulfilled precisely because the payload is constrained.

## The Intelligence Layer: Building on Semantic Data

The third layer is where developers build. Once Visual Distillation has converted the raw visual stream into structured semantic data, that data stream becomes the foundation for applications, and it's a better foundation than raw video in almost every respect.

Semantic data is faster to process, cheaper to store, and easier to transmit than raw footage, and a more consistent input to downstream models, because unlike raw frames it does not vary with network quality or frame degradation. Many of the performance and reliability problems we encountered at L+R while building on raw video capture, such as battery drain, degradation, and latency, improved when we reoriented our architecture around semantic data streams. The hardware constraints of wearables make raw video transfer uniquely punishing. Streaming raw footage over a Bluetooth connection creates massive bandwidth bottlenecks and rapid battery drain. Conversely, saving footage locally to the glasses for a later transfer requires a device-to-device Wi-Fi connection, which is equally power-hungry and destroys the real-time nature of ambient computing. By producing and transmitting lightweight semantic data instead, response times dropped because on-device inference removes the cloud round trip, power draw fell because the device no longer streams raw video over the radio, and the same semantic stream could feed several applications at once.

Semantic data is inherently shareable. Where raw footage acquisition requires a dedicated pipeline, with a single application capturing, processing, and holding exclusive access to a video stream, semantic data can be consumed simultaneously by multiple applications. A navigation aid, an accessibility tool, and a workflow assistant can all read from the same structured stream without any of them needing to independently acquire or process raw footage. This multi-threaded access is architecturally impossible in the Capture-Store and practically straightforward in Ephemeral Perception.

Developers building on semantic data can create context-aware AI assistants that understand a user's environment without storing a record of it. They can build memory systems that know what a user has encountered without retaining visual evidence of those encounters. They can build accessibility tools, workflow automation systems, spatial navigation aids, real-time translation layers, and intelligent retail experiences, all without ever generating a single frame of footage that could compromise the privacy of the wearer or the world around them.

| Dimension | Capture-Store | Ephemeral Perception |
|---|---|---|
| **Privacy model** | Policy-based. Raw footage exists and is protected by terms of service & consent language | Architecture-based. Raw footage never exists beyond the moment of capture and only on the user's device(s) |
| **Data created** | Raw video frames and audio data, stored and transmissible | Structured semantic data only. Labels, coordinates, text strings, presence signals |
| **Processing location** | Cloud server. Raw footage must travel off-device before intelligence is applied | On-device. Inference happens at the edge before any data leaves the user's device(s) |
| **Battery performance** | High drain. Continuous media encoding and the massive radio power (Bluetooth or Wi-Fi) required for wireless video streaming cause severe battery depletion. | Lower drain. Compute is distributed efficiently between the wearable's NPU and/or the tethered smartphone. This localized processing avoids heavy radio streaming, and syncing the resulting semantic data requires negligible power. |
| **Latency** | Higher. Cloud round trip adds processing delay | Lower. On-device inference eliminates network dependency |
| **Accuracy** | Variable. Dependent on network quality and cloud processing pipeline | More consistent. On-device inference is not subject to network degradation |
| **Regulatory exposure** | High. Raw visual data triggers biometric privacy laws, GDPR special category provisions, and BIPA | Low. Semantic data that cannot identify individuals falls outside most high-risk regulatory categories |
| **Enterprise adoption** | Blocked or slowed by legal and compliance review in sensitive environments | Accelerated. Architectural privacy guarantee removes primary procurement barrier |
| **Unintended leakage of visual data** | Inherent. Raw footage that travels to cloud infrastructure can be reviewed by human annotators, scraped by bad actors, etc. | Eliminated. No raw footage exists to be reviewed |
| **Training data utility** | Raw footage requires expensive, privacy-compromising human annotation to train AI models. | Base NPUs are trained offline on licensed data, while the resulting semantic output safely and instantly trains the user's personal AI. |
| **Reversibility** | Raw footage can be reconstructed, subpoenaed, leaked, or breached | Distillation is irreversible. Semantic output cannot be used to reconstruct source imagery |
| **Context awareness** | Potentially detached and requires adjacent data the vendor may not | Immediate access to data typically stored or cached on device, such as |

| Dimension | Capture-Store | Ephemeral Perception |
|---|---|---|
| | have access to (and ideally wouldn't have access to) | contacts, calendars, images, etc. Context becomes actionable on the edge. |
| **Security** | Increases the promptware attack surface[13] as raw data can be infected as it leaves the edge (or anywhere else in the inference pipeline). | Mostly eliminated, since the primary promptware attack surface is embedded in the edge device and inaccessible. Attack surfaces still exist for those inferences that require off-device inference. |

# IV. Ephemeral Perception in Practice: A Reference Application

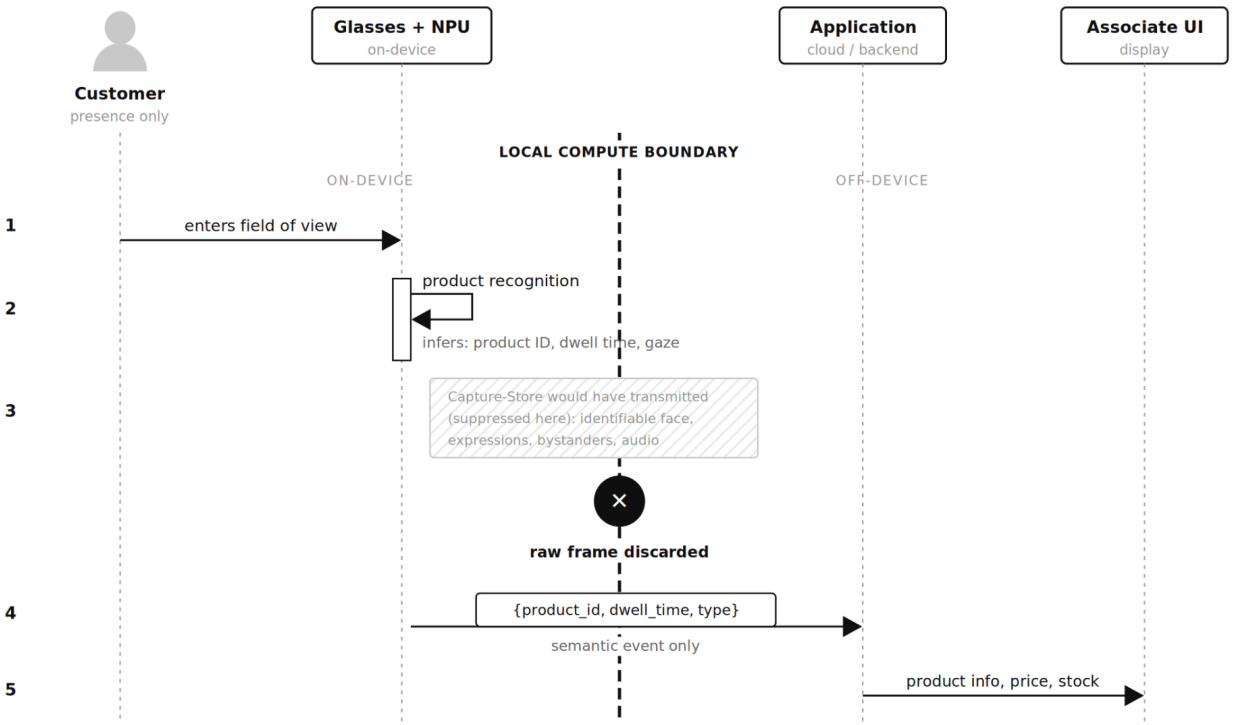
To make the framework concrete, consider how an Ephemeral Perception–compliant application might work in a real-world context.

Imagine a retail assistant application built on Meta's DAT for a luxury brand. The application's job is to help a sales associate understand what a customer is looking at, surface relevant product information in real time, and log customer interest signals for later analysis.

Under a Capture-Store architecture, the application would stream video and audio from the glasses to a cloud server, process it there to identify products and customer behavior, store that footage as part of the session record, and return results to the associate. The customer's face, expressions, behavior, and anyone's image caught in the background would be recorded and transmitted without their knowledge or consent.

Under Ephemeral Perception, the architecture is fundamentally different (Figure 4). The camera on the glasses captures the scene. On-device inference, running on the Snapdragon Wear Elite's neural processing unit, identifies the products the customer is looking at and generates a structured data event: product ID, dwell time, interaction type. That semantic event is transmitted to the application. The raw frame is discarded immediately and never leaves the device. No identifiable images are exposed. The associate receives the same useful information. The customer's privacy is preserved by the architecture of the system, not by a policy statement.

**FIGURE 4: The Retail Reference Application**
What the associate receives versus what is never created




The associate receives the same useful information. The customer's identity is never created, transmitted, or stored.

Within this architecture, the application produces better outcomes in every dimension: faster response times because on-device inference is lower latency than a cloud round trip, better battery performance because raw video is never transmitted, stronger privacy because no identifiable footage ever exists, and a cleaner compliance posture because the system architecturally cannot produce the kind of data that creates regulatory exposure.

This is not a hypothetical future capability. Every element of this architecture is buildable on the DAT today, though not every layer is yet turnkey, and many production systems will run a hybrid edge-cloud split in the near term.

# V. The Physical AI Paradox

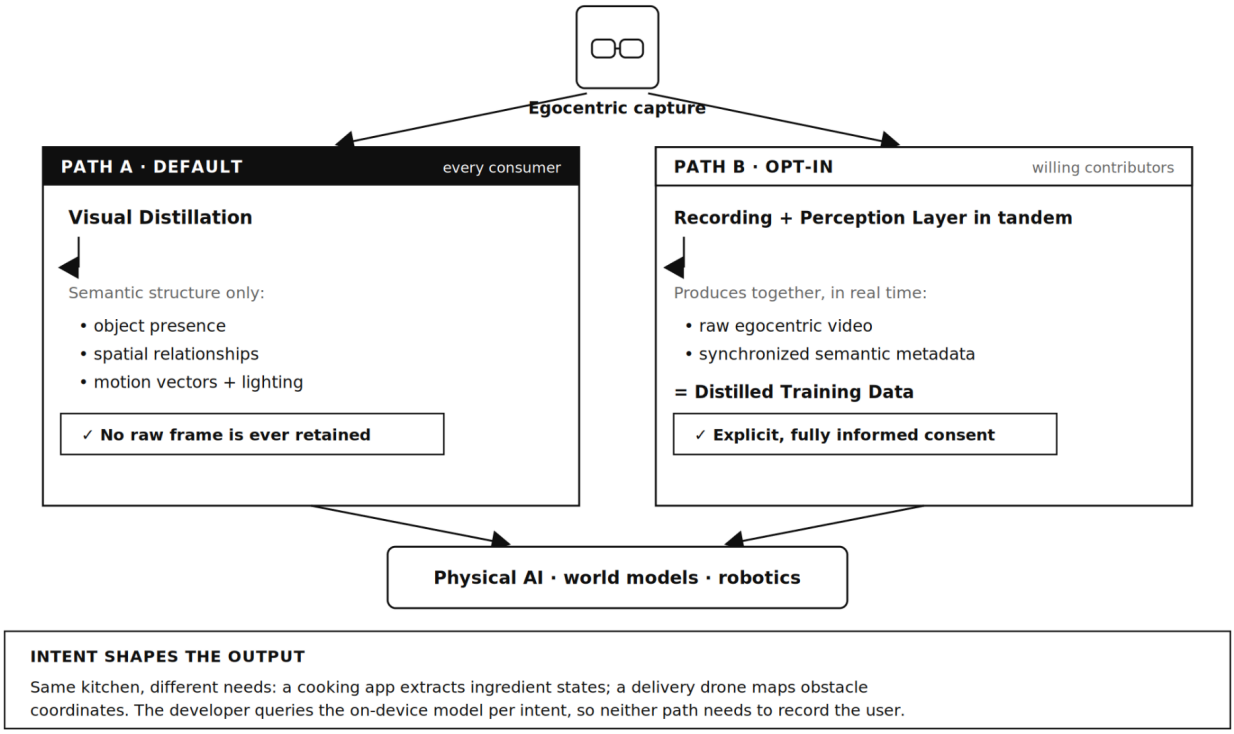
The stakes of getting this architecture right extend well beyond personal privacy. They reach into one of the most consequential technological developments of the coming decade: Physical AI.

Physical AI is the emerging field concerned with training AI systems to understand and operate in the physical world. Unlike the large language models that process text or the image generators that work with pixels, Physical AI systems need to learn how humans experience real environments: how we navigate spaces, interact with objects, read contexts, and make decisions in three dimensions.

The primary training signal for Physical AI is first-person visual data, often referred to as egocentric data, and the most abundant source of this egocentric data at scale is wearable devices.

This is where the Capture-Store and Physical AI collide in a way that demands serious attention. The robotics industry, the autonomous vehicle sector, the surgical AI field, and the industrial automation market all have an enormous appetite for the kind of real-world, egocentric data that ambient AI wearables generate. The temptation to feed that appetite by accumulating raw footage at scale is significant, and some of it is already happening.
The Physical AI Paradox is this: Physical AI needs to learn how humans experience the physical world, and the most direct path to that data runs through devices that millions of people wear on their faces in their most private moments. Collecting that data the way the Capture-Store collects it creates a surveillance infrastructure that will erode the public trust that ambient computing needs to achieve mainstream adoption. (This erosion has already started.)
Ephemeral Perception resolves the paradox by offering a dual path (Figure 5). For everyday consumer applications, Physical AI systems do not actually need to hoard raw ambient footage to learn. They need structured semantic data: object presence, spatial relationships, surface properties, motion vectors, lighting conditions, and interaction patterns. All of this can be derived through Visual Distillation without retaining a single frame of the user's private life. However, in scenarios where developers or fully aware users explicitly opt in to record themselves specifically to train robotics or AI models, the framework remains equally critical. Instead of merely transmitting raw video, the Perception Layer operates in tandem with the recording, generating rich semantic metadata in real time. This creates a more valuable training asset: raw egocentric video paired instantly with structured, machine-readable context, eliminating the separate annotation pass that raw footage would otherwise require.

**FIGURE 5: One Sensor, Two Paths**
How Ephemeral Perception serves everyday use and opt-in model training from the same capture




A robot learning to navigate a kitchen is a useful example. For a consumer user, the robot does not need footage of them cooking; it only needs the semantic structure of the environment. Crucially, semantics depend on intent. A cooking assistant and a delivery drone require different knowledge about the same kitchen. Ephemeral Perception lets developers query the on-device model for specific needs—extracting ingredient states or mapping obstacle locations—without recording the user.

Whether operating in a strict privacy mode for consumers or a metadata-generating mode for willing contributors, the emerging industry term for this approach is Distilled Training Data. In many cases, it provides a superior training signal: cleaner, more consistently labeled, and easier to scale than raw footage, which often requires extensive human annotation before use.

The data annotators reviewing intimate footage in the Meta and Sama investigation were performing a step that Visual Distillation either heavily reduces or makes entirely unnecessary.

# VI. The Platform Explosion: Why a Shared Standard Matters Now

The argument for Ephemeral Perception would be compelling even if only one platform were involved. The argument becomes urgent when you look at the full scope of what's launching in 2026.

The ambient computing market currently lacks a single, unified platform where one corporate policy can govern data collection. Instead, the hardware landscape is rapidly diversifying. Meta operates a tightly controlled developer ecosystem through its Device Access Toolkit. Apple enforces its own strict, walled-garden privacy architecture across its spatial computing and wearable lines.

Google confirmed Android XR, its open platform for smart glasses, and used Google I/O 2026 to detail the full developer roadmap, including expanded SDK access, deeper Gemini integration for on-device perception, and a developer toolchain designed to span everything from lightweight AI glasses to full mixed-reality headsets. Samsung began quietly laying the software groundwork for its upcoming 'Galaxy Glasses' in its latest developer betas. The platform is designed to power an extensive range of form factors. These include video see-through headsets like the Samsung Galaxy XR, which is positioned to compete directly with the Apple Vision Pro and Meta Quest Pro, as well as optical see-through headsets, augmented reality glasses, and lightweight AI glasses. Beyond traditional tech manufacturers, this open ecosystem has expanded to include global fashion and eyewear conglomerates. Brands such as Warby Parker and Gentle Monster are integrating these technologies, and Kering Eyewear recently announced a major partnership with Google to develop AI-powered luxury glasses.[14] By seamlessly blending high-end design with Android XR's contextual awareness, these partnerships ensure that ambient computing will reach a massive, fashion-conscious consumer base.

Snap occupies a distinct position. Through Spectacles and its Snap OS platform, the company has built one of the most mature AR developer ecosystems in the market, with hundreds of thousands of Lens creators, and has confirmed a consumer launch of its next-generation Specs in 2026.[15]

Simultaneously, the market is flooding with independent challengers and international players. Companies including Alibaba, Brilliant Labs, XREAL, Even Realities, Lucyd, Mentra, Rokid, and SOLOS are introducing ambient wearables governed by entirely different regulatory norms. Furthermore, supply chain reports from April 2026 indicate that OpenAI is actively co-developing its own AI-native smartphone with Qualcomm and MediaTek.[16] This move is

designed to bypass traditional app ecosystems, giving the leading AI developer total control over the operating system and real-time device sensors to power autonomous AI agents. With so many disparate players entering the physical hardware space, these devices are bound by no shared standards at all.

Below this layer of named challengers sits a long tail that compounds the fragmentation problem. A search for 'smart glasses' on Amazon today returns thousands of products ranging from roughly seven-dollar Bluetooth audio frames with no camera to eighteen-hundred-dollar XR display glasses with micro-OLED panels and integrated AI assistants. Many of the camera-equipped devices in this long tail are produced by manufacturers with no consumer-facing privacy reputation to protect, and many route their visual and voice processing through shared third-party companion apps. One such companion app, HeyCyan, published by Shenzhen Qingcheng Future Technology Co., Ltd., serves as the AI backend for a range of Chinese-manufactured smart glasses and routes user voice and vision queries through large foundation models including OpenAI's GPT-4o and Alibaba's Tongyi Qwen, depending on geography and feature.[17] The user experience is a friendly mobile app. The underlying data architecture is Capture-Store, executed by parties most users will never know exist.

This fragmentation demonstrates why a shared architectural framework matters more than platform-specific privacy policies. Relying on app store reviews, corporate terms of service, or updated consent forms to protect the public's visual reality is an inadequate strategy for an open ecosystem. An architectural default that destroys raw data before it becomes a transmissible liability provides a much more reliable defense.

The long-term consequences of unstandardized data hoarding are already visible. As reported by the MIT Technology Review, the photographic content generated by hundreds of millions of users on Pokémon Go is now being used by a Niantic spinoff to build a world model for last-mile delivery robots, overcoming the limitations of GPS data in urban settings.[18] While this clearly advances urban robotics, there is little transparency regarding how that historical data has been processed and treated. The public cannot verify if Personally Identifiable Information (PII) is still part of those datasets, nor is the exposure risk clear if the company were to suffer a cyberattack.

The window to establish Ephemeral Perception as the default approach remains open. The developer community building on these emerging platforms is still small enough to be shaped by a compelling technical argument. This first generation of ambient AI applications will set the norms that subsequent generations inherit. If raw visual capture becomes the default standard today, it will calcify into infrastructure and business models that become increasingly difficult to reverse.

# VII. What Developers Should Do Now

The framework described in this paper is philosophical and architectural, but its implications are practical and immediate. For developers building on Meta's DAT, Google's Android XR, or any other ambient AI platform in 2026, here are the concrete principles to build around. These five principles distill the architectural posture into actionable guidance.

## 1. Treat every frame as a liability until it becomes data

The default mental model in most development pipelines treats raw visual data as a resource to be collected and stored for later use. Ephemeral Perception inverts that model. Every raw frame that exists beyond the moment of inference is a liability: a privacy risk, a storage cost, a regulatory exposure, and a performance bottleneck. The goal of every processing decision should be to convert that liability into an asset, structured semantic data, as quickly as possible and discard the source material immediately.

In practice this means designing your processing pipeline so that inference happens first, before any storage or transmission decision is made. The frame is processed, the meaning is extracted, and the frame is discarded. That sequence should be enforced at the architecture level, not left to developer discipline or policy.

## 2. Define your semantic schema before you write a line of code

Before building any application on a visual data stream, define precisely what structured data your application actually needs. Not what you might need, not what could be useful later, but what your application requires to deliver its core value. That schema, the specific labels, coordinates, text strings, presence signals, and action events your application depends on, is the only data your system should ever produce or retain. Everything else is noise that creates risk without creating value.

This discipline also produces better applications. A tightly defined semantic schema forces clarity about what the application actually does, which produces faster processing, more reliable outputs, and a simpler, more maintainable codebase.

## 3. Push inference to the edge wherever the hardware allows

The Qualcomm Snapdragon Wear Elite and the broader trend toward on-device AI processing means that edge inference is no longer a compromise. For many common computer vision tasks, running inference on the device produces lower latency, better battery performance, and

stronger privacy guarantees than transmitting to the cloud. Evaluate your inference requirements against current on-device capabilities before defaulting to cloud processing. The performance gap that once made cloud processing the only viable option for sophisticated vision tasks is narrowing rapidly and for many use cases has already closed.

## 4. Make Consent-by-Architecture your compliance strategy

Policy-based privacy compliance, relying on consent flows, terms of service language, and user-facing controls, is increasingly inadequate as a sole strategy for ambient AI applications. Regulators and courts are moving toward evaluating not just what policies say but what systems actually do. An application that architecturally cannot produce identifiable visual records of anyone, because Visual Distillation discards the raw signal before it ever leaves the device, has a fundamentally stronger compliance posture than one whose privacy protection depends entirely on a user reading and understanding consent language and the developer code actually conforming to that language.

Build your compliance argument around what your system cannot do, not just what your policy says it won't do. That distinction will matter increasingly as regulatory scrutiny of ambient AI wearables intensifies.

## 5. Document your data lifecycle explicitly

For every category of data your application touches, be able to answer four questions: where does it come from, what happens to it at the point of capture, where does it go, and when is it destroyed. If you can't answer all four questions clearly for every data category in your application, your data lifecycle isn't defined well enough. Publish that documentation, either in your application's privacy policy or as a standalone technical document. The developers who can demonstrate a clear, auditable data lifecycle will have a significant advantage as enterprise procurement, regulatory review, and consumer trust all increasingly demand it.

# VIII. The Regulatory Horizon: Building Ahead of the Curve

Regulators in the United States, European Union, and United Kingdom are watching the ambient AI wearables space with increasing attention. The direction of travel across all three jurisdictions is toward stricter requirements for biometric and visual data handling. Legal scholars tracking these frameworks consistently identify ambient computing as one of the highest-risk categories for regulatory action in the near term. Understanding where that regulatory pressure is coming from, and where it is heading, is essential for developers and enterprises making architecture decisions today.

## GDPR and the EU AI Act

The European Union's General Data Protection Regulation (GDPR) treats biometric data, which includes facial images and other visually derived personal identifiers, as a special category requiring explicit consent and strict processing limitations.[19] Ambient AI wearables that capture and transmit raw visual data in European jurisdictions face significant exposure under these provisions, particularly when the footage includes bystanders who have not consented to being captured.

The EU AI Act, which entered into force in 2024 and is being phased in through 2026 and beyond, establishes a risk-based framework for AI systems. It places real-time biometric identification systems in the highest risk category, subjecting them to strict requirements and in many cases outright prohibition in public spaces.[20] Applications built on Ephemeral Perception principles, which by architecture cannot produce real-time biometric identification of individuals, are positioned significantly better than applications built on raw visual capture. The pace of platform expansion makes this regulatory posture increasingly consequential. Between Meta Connect 2025, Google I/O 2026, and the CES/MWC 2026 announcements, the developer ecosystem onboarding to ambient computing has expanded by an order of magnitude in under twelve months, without any corresponding harmonization of how raw visual data is to be handled across platforms.

## US State Privacy Laws and the UK ICO

In the United States, Illinois' Biometric Information Privacy Act (BIPA) remains the most consequential state-level regulation for ambient AI applications. Its requirement of informed written consent before collecting biometric identifiers, paired with a private right of action, has already produced substantial settlements against major technology companies.[21] California's

CPRA,[22] Texas' CUBI Act,[23] and a growing roster of state biometric privacy laws are widening this landscape quickly. Federal privacy legislation, while not yet enacted, is advancing further than at any prior point, with biometric data consistently named as a priority area. The architectural advantage is the same one that holds under GDPR: an application that cannot generate a biometric identifier in the first place sidesteps the core trigger these statutes are built around.

In the United Kingdom, the Information Commissioner's Office called the practices surfaced in the Swedish investigation "concerning" and said it would press Meta on its compliance with UK data protection law, an early signal that the ambient AI category is now squarely on the regulator's radar.[24]

# IX. Conclusion: The Architecture of Trust

The ambient computing era has arrived not with a single announcement but with an accumulation of moments: a Swedish newspaper investigation into footage no one expected to be watched, a chip announcement in Barcelona that makes cloud processing optional for the first time, a platform opening its camera API to outside developers for the first time, and dozens of devices shipping to millions of people who will wear them in the most ordinary and intimate moments of their lives.

Each of these moments individually is a technology story. Together they describe a seismic shift in how human beings relate to computation and a narrowing window in which the foundational norms of that shift can be established.

The argument this paper has made is simple at its core. The dominant architecture for ambient AI wearables, built around capturing, transmitting, and storing raw visual data, was never the technically superior default. It was the only viable option in an earlier era of connected devices, when on-device processing was too limited to support the alternative. That limitation no longer exists. The hardware has arrived. The computer vision systems capable of distilling raw visual input into structured semantic meaning at the point of capture, without ever generating a transmissible visual record, are real and available now. The silicon to run them on device, without a cloud round trip, is shipping now.

What remains is a question of industry norms and pragmatic execution. The temptation for engineering teams will always be to fall back on familiar image processing libraries and legacy cloud architectures simply to meet product shipping deadlines. Asking developers to learn how to build around semantic extraction rather than scanning raw pixels introduces a significant learning curve. Will the developer community building the first generation of ambient AI

applications surrender to that friction and inherit the default architecture with all its liabilities? Or will they recognize that mastering Ephemeral Perception today is the only sustainable way to build for tomorrow?

This white paper contends that moving to Ephemeral Perception now, while a steep learning curve, is a smaller cost in the long run than starting with Capture-Store and later being forced to migrate, paying a second implementation cost while also managing the litigation that Capture-Store will inevitably produce.

Moving to Ephemeral Perception now is not a regulatory requirement or a constraint imposed from the outside. It is an architectural philosophy that pays off across the dimensions that matter most for ambient deployment: lower latency and lower power draw from keeping inference on-device, behavior that does not degrade with network conditions, a compliance posture built on data the system never generates, faster enterprise procurement, and the basic dignity of every person who will appear in the field of view of a device they never chose to be near.

The person in the bathroom, the colleague in the meeting, the customer in the store, the stranger on the street: none of them signed up to be part of an AI training dataset. The architecture of the devices worn by the people around them will determine whether they are. That architecture is being written right now by developers making decisions that feel small and technical but are anything but.

We publish this framework as a reference architecture because we believe the ambient computing ecosystem will be stronger if its foundational norms are established by practitioners who have confronted these tradeoffs in real products, rather than by regulators who arrive after the damage is done. We invite every developer, hardware manufacturer, platform owner, and policymaker who reads this to adopt, adapt, and build on what we have proposed here.

The camera was never the point. The meaning was always the point. Build for the meaning, and everything else follows.

# Glossary

**Ambient Computing** — A paradigm where technology operates passively in the background through wearables and sensors, shifting interaction from active screen engagement to natural, contextual awareness.

**Biometric Data** — Personal data derived from physical or behavioral characteristics, such as facial geometry. In ambient computing, raw visual data frequently contains biometrics, triggering strict regulatory frameworks like GDPR and BIPA.

**Capture-Store** — The legacy data architecture of the modern digital economy. It relies on capturing raw user data, transmitting it to the cloud for processing and storage, and returning a result. In ambient computing, it creates an unavoidable, high-risk chain of custody for continuous video.

**Consent-by-Architecture** — Privacy guaranteed by system design rather than legal enforcement. Ephemeral Perception is the canonical example: by processing data locally and discarding raw inputs instantly, data misuse and regulatory breaches become technically impossible at the code layer, not merely prohibited by policy.

**Distilled Training Data** — Structured semantic data derived from on-device computer vision. It trains spatial models without retaining raw visual records or personal information.
Egocentric Data — First-person visual, audio, and spatial data captured from the user's perspective. It is the highly valuable but privacy-sensitive raw material traditionally used to train spatial computing systems.

**Ephemeral Perception** — The core philosophy that ambient systems should perceive the world to extract meaning, but must discard raw sensory data immediately and irreversibly at the exact point of capture.

**On-Device Inference** — The execution of machine learning models directly on the local hardware (such as a wearable device) without transmitting the source data to cloud servers.
Perceptual Sovereignty — An architectural guarantee that raw sensory data never exits the local compute boundary of the device.

**Personally Identifiable Information (PII)** — Any data that could potentially identify a specific individual. In the context of ambient wearables, raw video feeds are inherently dense with PII.
Physical AI — Artificial intelligence focused on understanding and operating within the physical world, relying heavily on first-person spatial and visual data for training.

**Prompt Injection** — The foundational exploit of the generative AI era. It involves inserting malicious data (e.g., text, pixels, or instructions) into an AI system's input stream to override its original system instructions and hijack its behavior. One of the first examples showed how inserting malicious pixels into a picture caused the AI system to identify a stop sign when shown a picture of a yellow school bus.

**Promptware** — Software architectures built fundamentally around large language models and generative AI. In these systems, natural language or multimodal sensory inputs act as executable code rather than static data, radically expanding the traditional attack surface.

**Semantic Payload** — The structured data that remains after Visual Distillation, such as labels, spatial coordinates, and relational graphs. It provides complete developer utility while remaining private by construction.

**Spatial Computing** — The digitization of activities, objects, and relationships in 3D physical space, enabling digital information to interact with the real world natively.

**The Physical AI Paradox** — The tension between Physical AI's need for massive amounts of first-person spatial data and the privacy violations caused by collecting it via traditional Capture-Store methods. Resolved through Distilled Training Data.

**Visual Distillation** — The one-way process of converting raw visual input into semantic data. Like a one-way function, it runs forward to extract meaning but can never be reversed to recreate the original image.

**World Model** — An AI system's internal spatial representation of the physical environment, mapped and updated using visual data to help the system understand physics, depth, and object relationships.

# Endnotes


1. Vincent Liu, Ademi Adeniji, Haotian Zhan, Raunaq Bhirangi, Pieter Abbeel, Lerrel Pinto: "EgoZero: Robot Learning from Smart Glasses." arXiv preprint arXiv:2505.20290, May 2025. [https://arxiv.org/abs/2505.20290]

2. "VisionClaw: Always-On AI Agents through Smart Glasses." arXiv preprint, submitted by researchers Xiaoan Liu, DaeHo Lee, Eric J Gonzalez, Mar Gonzalez-Franco, and Ryo Suzuki, April 2026. [https://www.researchgate.net/publication/403561359_VisionClaw_Always-On_AI_Agents_through_Smart_Glasses/download]

3. Meta Connect 2025 keynote and official announcement, September 18, 2025. Meta for Developers Blog: "Introducing the Meta Wearables Device Access Toolkit," officially announcing early partners including Disney Imagineering, Logitech, Microsoft, and L+R. [developers.meta.com/blog/introducing-meta-wearables-device-access-toolkit/]

4. Google I/O 2026 keynote and Android XR developer announcements, May 2026. "What's new in Android XR at I/O 2026." [https://blog.google/products-and-platforms/platforms/android/android-xr-io-2026/]

5. CES 2026, Las Vegas. Multiple AI wearable device announcements including AI pins, ambient pendants, smart glasses, and connected rings across dozens of manufacturers.

6. Mobile World Congress 2026, Barcelona. Qualcomm Snapdragon Wear Elite announcement and multiple AI wearable manufacturer announcements.

7. Alibaba global availability announcement for Qwen AI Glasses, 2026. [https://baike.baidu.com/en/item/Qwen%20AI%20Glasses/1469262]

8. Smart Analytics Global (SAG), "AI Smart Glasses to Quadruple Revenue in 2026 as Apple and Samsung Prepare to Enter the Market," January 13, 2026. Revenue projected to quadruple to $5.6 billion by the end of 2026 (up from $1.2 billion in 2025); global shipments projected to rise from 6 million to 20 million units. [businesswire.com/news/home/20260113778367/en/AI-Smart-Glasses-to-Quadruple-Revenue-in-2026-as-Apple-and-Samsung-Prepare-to-Enter-the-Market-Says-Smart-Analytics-Global-SAG]

9. Svenska Dagbladet and Göteborgs-Posten, joint investigation into Meta Ray-Ban smart glasses data annotation practices, February 2026.

[https://www.svd.se/a/K8nrV4/metas-ai-smart-glasses-and-data-privacy-concerns-workers-say-we-see-everything]

10. BBC News, coverage of the Svenska Dagbladet and Göteborgs-Posten investigation, March 2026. [https://www.bbc.com/news/articles/c0q33nvj0qpo]

11. Meta AI research announcement, November 2025. "SAM 3: Segment Anything Model 3." [https://ai.meta.com/research/sam3/]

12. Qualcomm press release, Mobile World Congress 2026. Snapdragon Wear Elite announcement with dedicated Hexagon NPU for wearable devices. [https://www.qualcomm.com/snapdragon/news/unveiled-at-mwc-2026--truly-personal-ai-powered-by-snapdragon-we]

13. Oleg Brodt, Elad Feldman, Bruce Schneier, Ben Nassi: "The Promptware Kill Chain: How Prompt Injections Gradually Evolved Into a Multistep Malware Delivery Mechanism," February 2026. [https://arxiv.org/abs/2601.09625]

14. Kering Eyewear press release, 2025. "Kering Eyewear partners with Google to develop smart glasses." [https://keringeyewear.com/newsroom/2025/Kering-Eyewear-partners-with-Google-to-develop-smart-glasses]

15. Snap Inc., "Snap to Launch New Lightweight, Immersive Specs in 2026," Augmented World Expo, June 10, 2025 (describing Snap OS, the Remote Service Gateway for camera access, and integrations with OpenAI and Gemini); and Specs Inc. (a Snap subsidiary) and Qualcomm, "Snap and Qualcomm Expand Strategic Collaboration to Advance Intelligent Computing Experiences on Specs," April 10, 2026. [https://newsroom.snap.com/launch-specs-2026] [https://newsroom.snap.com/snap-qualcomm-strategic-collaboration-specs-2026]

16. Ming-Chi Kuo, TF International Securities. Supply chain industry report regarding OpenAI smartphone processor development with Qualcomm, MediaTek, and Luxshare, April 27, 2026.

17. HeyCyan companion application, Shenzhen Qingcheng Future Technology Co., Ltd. Product documentation describing voice and vision routing through OpenAI's GPT-4o and Alibaba's Tongyi Qwen foundation models. [https://apps.apple.com/us/app/heycyan/id6742974094] [https://heycyan.net/]

18. MIT Technology Review, "How Pokémon Go is giving delivery robots an inch-perfect view of the world," March 2026. [https://www.technologyreview.com/2026/03/10/1134099/how-pokemon-go-is-helping-robots-deliver-pizza-on-time/]

19. European Union General Data Protection Regulation (GDPR), in effect May 2018.

20. European Union Artificial Intelligence Act, entered into force August 2024, phased implementation through 2026 and beyond.

21. Illinois Biometric Information Privacy Act (BIPA), 740 ILCS 14, enacted 2008.

22. California Consumer Privacy Act (CCPA), as amended by the California Privacy Rights Act (CPRA), effective January 2023.

23. Texas Capturing or Use of Biometric Identifier Act (CUBI), Texas Business and Commerce Code Chapter 503.

24. BBC News, 2026. The UK Information Commissioner's Office (ICO) stated the reported practices were "concerning" and that it would write to Meta to request information on its compliance with UK data protection law. [https://www.bbc.com/news/articles/c0q33nvj0qpo]