

Visual Context for Everyday Wearables: A Literature Review of Smart Eyewear, Technology Acceptance, Bystander Privacy and Modular Sensing

Dr. A. B. Jirapure¹, Rutusha Somkuwar², Aryan Wadibhasme³, Deepak Sharma⁴, Yash Ingole⁵

¹Professor & Head of Department, Department of Industrial IoT, Priyadarshini College of Engineering, Nagpur, India

^{2,3,4,5}UG Scholars, Department of Industrial IoT, Priyadarshini College of Engineering, Nagpur, India
Affiliated to Rashtrasant Tukdoji Maharaj Nagpur University, Nagpur - 440019, Maharashtra, India

Abstract - Smart glasses spent more than a decade moving from showy consumer demonstrations to real mass-market sales, and researchers still disagree about what held them back. This narrative review follows head-worn computing from the first head-mounted displays to camera-equipped glasses paired with multimodal AI assistants. Along the way it draws together work on technology acceptance, bystander privacy, vision-language models, edge-cloud processing and human-centric Industry 5.0 manufacturing. Scholarly sources came from IEEE, ACM, Springer, Elsevier, JMIR and arXiv, extended by citation tracing and by primary market releases from the International Data Corporation (IDC). Three conclusions emerge. People adopt smart glasses when the frames look socially acceptable and the device is useful from the first day, while bystander privacy stands as a separate barrier whose weight shifts with the setting. The revival since 2023 coincides with multimodal conversational models that gave a head-worn camera a daily purpose, although one ecosystem dominates the market. And inexpensive eyewear has largely meant audio without a camera. Significant gaps remain. We found no peer-reviewed study of modular visual assistance assembled from a commodity microcontroller, a smartphone and existing true-wireless stereo (TWS) earbuds. Few benchmarks cover single low-resolution snapshots taken in poor light, privacy safeguards remain laboratory prototypes, and end-to-end latency, energy use and accuracy have not been measured. The review closes with a five-part roadmap covering willingness to pay, bystander-inclusive acceptance, latency-accuracy benchmarking, privacy filtering and consented shop-floor pilots.

Key Words: Smart glasses, Technology acceptance, Bystander privacy, Multimodal AI, Wearable computing, Edge computing, Industry 5.0, Literature review

1. INTRODUCTION

Computing worn on the head has been a goal for more than fifty years. The idea began with early head-mounted displays and grew into the wearable-computing vision of a machine worn all day that extends its owner's memory and senses. Google Glass carried head-worn computing to

the public in 2013, but its reception was rocky and the product did not last commercially. Consumer adoption at volume shows up in industry trackers only in the mid-2020s: IDC recorded about 2.25 million display-less smart glasses shipped worldwide in the first quarter of 2026, a 167% rise on the previous year [1].

That jump invites a simple reading, that the hardware finally grew up. The research tells a messier story. Smart eyewear has always lived with a tension between what the technology can do, whether people can wear it comfortably, and whether those around the wearer will tolerate it, and no single hardware breakthrough has resolved that tension. We therefore pose three questions:

RQ1: Why did consumer acceptance lag behind hardware availability for more than a decade, and which behavioral explanations hold up against the evidence?

RQ2: How did generative multimodal artificial intelligence (AI) change what a head-worn camera is for?

RQ3: What separates cheap audio-only eyewear from premium camera-equipped AI glasses, especially in price-sensitive markets such as India, and could visual contextual intelligence be made affordable through a decoupled, modular service layer?

This is a narrative review. We built no prototype and ran no experiments. Instead we place several bodies of literature next to one another, weigh the strength of their evidence, and note where that evidence stops. Section 2 describes the method, Section 3 traces the evolution of smart eyewear, Section 4 covers technology acceptance and privacy, Section 5 turns to market evidence, Section 6 to multimodal edge computing, and Section 7 to Industry 5.0. Section 8 compares the studies reviewed, Section 9 sets out gaps and a five-part agenda, and Section 10 concludes.

2. REVIEW METHODOLOGY

We chose a narrative design over a registered systematic protocol such as PRISMA for two reasons. The topic

crosses fields that seldom cite one another (human-computer interaction, information systems, computer vision, mobile distributed systems, legal studies and industrial engineering), and much of the key market data sits outside bibliographic indexes.

Scholarly sources came mainly from IEEE, ACM, Springer, Elsevier, JMIR and arXiv. The search began with seminal papers on wearable computing and technology acceptance and then widened by following citations backwards and forwards (snowballing). Market figures were taken only from primary IDC releases, read on IDC's own portals in October 2026; unverified press coverage and paywalled secondary summaries were left out. One open-source maker project was included, solely to document which microcontroller-based baselines already exist.

We kept studies that were peer reviewed, or were authoritative monographs, standards or legislation, and that spoke directly to at least one research question. Marketing whitepapers, undated blogs, anonymous commentary and product reviews without technical specifications were dropped. Where a DOI existed, reference details were checked against the publisher's record.

Three limitations should be kept in mind. The selection was purposive rather than exhaustive, and study quality was not scored numerically. Market categories change from one tracker quarter to the next, so the figures cannot be pooled statistically. And cloud API prices and platform policies move quickly, so any cost figure here is a snapshot in time.

3. EVOLUTION OF SMART EYEWEAR

3.1 Foundations and the first consumer wave

By smart glasses we mean spectacle-mounted wearables that combine environmental sensing, onboard computing and wireless connectivity with audio or visual output. The first consumer wave, best known through the Google Glass Explorer Edition, met engineering and social obstacles at the same time, and the social ones proved more damaging. In situ studies found that bystanders were deeply uneasy about covert video capture and wanted visible recording cues and a clear way to say no [2].

3.2 Audio-centered eyewear

The second phase, roughly 2017 to 2021, dropped the display and the front camera and moved the computing into open-ear speakers in the temples. Removing the camera eased surveillance worries and saved weight. It also removed much of the reason to own the device for anyone who already had good wireless earbuds.

That critique is not universally accepted. Open-ear sound keeps the wearer aware of the surroundings and avoids ear-canal fatigue, and IDC India analysts expect open-ear wearables to grow as buyers put auditory comfort first [3]. Still, the safest conclusion is that audio-only glasses fill an ergonomic niche without adding any computing ability that earbuds lack. The three phases described here are our own analytical framework, not official industry categories.

3.3 Camera plus multimodal assistant

The current phase pairs normal-looking frames with image sensors, microphones and cloud-hosted multimodal foundation models. The change came from software rather than optics. Large multimodal models that read images and spoken questions together [4] gave the head-worn camera a daily use that a phone cannot match for convenience: answering questions about whatever the wearer is looking at.

The market numbers show growth that is fast but very uneven. After the 167% jump in display-less eyewear in Q1 2026 [1], overall XR hardware growth eased to 35.3% in Q2, and one vendor alone held 68.7% of global shipments [5]. Across all three phases the same pattern repeats: each generation solved one barrier and exposed the next. Fig. 1 shows the progression.

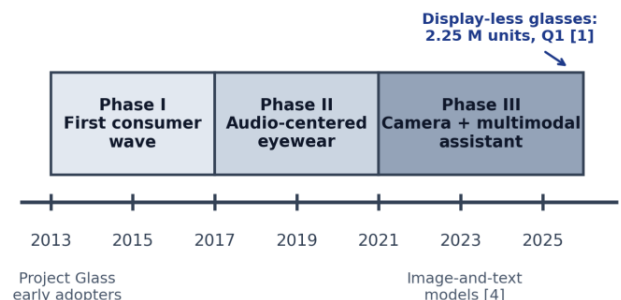


Fig -1: Phases of smart-eyewear development and milestones cited in this review (phase boundaries are analytical).

4. TECHNOLOGY ACCEPTANCE AND BYSTANDER PRIVACY

4.1 Acceptance models applied to wearables

If the first consumer wave stumbled on social grounds, the behavioral literature helps explain why. Research on wearable adoption leans on a handful of well-tested frameworks. The Technology Acceptance Model (TAM) holds that intention to use is driven mainly by perceived usefulness (PU) and perceived ease of use (PEOU) [6]. Later versions added social influence and job relevance (TAM2) and then the unified theory (UTAUT), which combines performance expectancy, effort expectancy,

social influence and facilitating conditions. UTAUT2 went further, adding hedonic motivation, price value and habit so that it could describe voluntary consumer choices [7]. For augmented-reality smart glasses, Rauschnabel et al. extended these models with privacy constructs [8].

All of these models share a weakness: they treat adoption as a private decision made by one person. That holds well enough for a smartphone or a smartwatch. It fails for a camera on someone's face, because the device changes how everyone nearby experiences the room. Fig. 2 shows a composite framework that adds an explicit bystander-privacy path to the classic constructs. It is a conceptual synthesis and has not been tested as a structural equation model.

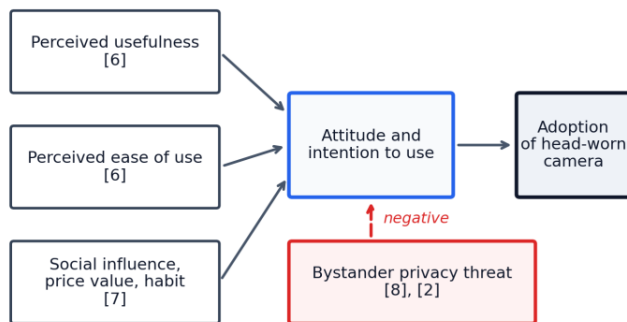


Fig -2: Composite acceptance framework for head-worn cameras (conceptual synthesis of [2], [6], [7], [8]).

4.2 Bystander privacy and social acceptability

Rauschnabel, He and Ro showed that the privacy threat a wearer poses to other people is an adoption barrier in its own right, distinct from the wearer's worries about personal data [8]. The reasoning is straightforward. Raising a phone signals that recording is under way and shows what is being pointed at. A camera built into an everyday frame gives no such signal. In field interviews, Denning et al. found that bystanders consistently wanted a clear physical sign of active recording and a practical way to refuse [2].

Technical remedies exist, but almost all remain laboratory prototypes. PrivacEye, for example, combines eye tracking with analysis of the first-person scene to close a mechanical shutter over the camera in sensitive social situations. It was tested on only 17 participants under controlled conditions [9].

How strongly bystander privacy shapes adoption across institutions is still unsettled. A structural equation modeling study of health-care professionals found that social-interaction constructs improved the model, yet expected privacy barriers were lower in clinics than

earlier studies of the general public had suggested [10]. One plausible reading is that patients and clinicians accept monitoring tools when they improve care. Factory floors may show similar tolerance, but we found no peer-reviewed study of shop-floor acceptance in Indian manufacturing.

5. MARKET EVIDENCE AND PRODUCT LANDSCAPE

Behavior alone does not explain the revival, so the market data deserve a closer look. They are untidy, because tracking agencies keep redrawing their categories. IDC separates display-less smart glasses from immersive optical-display XR headsets [1], [5]. Table 1 lists the market indicators used here, all taken from primary IDC releases.

Table -1: Market indicators from primary IDC releases

Indicator	Value	Source
Display-less glasses, Q1 2026	≈2.25 M units; +167% YoY	[1]
Display-less forecast	13.6 M (2026); 27.3 M (2030); CAGR 18.9%	[1]
XR shipment growth, Q1 and Q2 2026	130.2%; 35.3%	[5]
Leading vendor share, Q2 2026	68.7%	[5]
India wearables, Q2 2026	24.7 M units; -7.4% YoY	[3]
India earwear, Q2 2026	18.05 M units	[3]
India smart glasses, Q2 2026	69,400 units; +38.3% YoY	[3]

India looks different from North America and Europe. The wider Indian wearable market shrank 7.4% year on year in Q2 2026 (24.7 million units), while smart glasses grew 38.3% from a very small base of 69,400 units [3]. Smart glasses made up under 0.4% of the 18.05 million earwear units shipped that quarter [3].

This gap matters for our argument. Huge numbers of Indians already own personal TWS earwear, which is the precondition for any modular design that separates visual capture from audio playback. IDC analysts also expect earwear to become multifunctional, possibly with camera-enabled earbuds [3]. Such a shift would put competitive pressure on external smart-glasses modules, and that is a reason to test low-cost modular sensing soon. Fig. 3 plots Indian wearable shipments by category on a log scale.

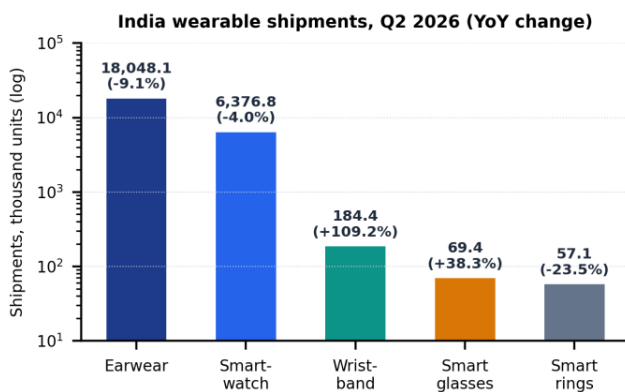


Fig -3: India wearable shipments by category, Q2 2026 (data from [3]).

Table 2 sorts current product types by what they sense and output, their trade-offs and their main limitation. The only publicly documented open-source design that fully separates the visual sensor from the audio output and the cloud AI backend is OpenGlass, a maker prototype built on an ESP32-S3 board with a claimed bill of materials under USD 25 [11]. It requires 3D printing, custom firmware flashing and user-supplied API keys, which confines it to hobbyists, and its repository states that primary software development has stopped.

Table -2: Product classes in the smart-eyewear space (qualitative synthesis)

Class	Sensing output /	Principal limitation	Basis
Open-ear audio eyewear	Microphones, speakers; no camera	Audio overlaps earbuds; no visual context	[3]; authors' assessment
AI camera glasses	Camera, microphones, cloud assistant	Higher entry price; one vendor dominant	[1], [5]
Display glasses	Display, often camera	Weight, power, price	[5]
Maker kits	Microcontroller camera, phone, third-party models	Assembly effort; self-managed software	[11]

Whether audio-only glasses can replace personal wireless earbuds has never been tested empirically. The peer-reviewed literature we found contains no rigorous acoustic benchmarking, battery testing or longitudinal preference study comparing the two. Qualitatively, open-ear glasses keep the wearer aware of the surroundings and can carry prescription lenses, while in-ear earbuds offer better noise isolation, sound quality and active noise cancellation. Our claim is therefore narrower than "glasses replace earbuds." Audio alone cannot justify a higher price

for glasses unless the frame adds something earbuds physically cannot: egocentric visual context.

6. MULTIMODAL AI, EGOCENTRIC VISION AND EDGE COMPUTING

6.1 From task-specific recognizers to conversation

Behind the market story sits a technical one. Early wearable cognitive assistance relied on rigid, task-specific vision classifiers. The Gabriel system of Ha et al. showed that real-time help with physical assembly tasks was possible within tight latency limits, but it depended on purpose-built cognitive engines running on nearby servers [12].

Transformer-based vision-language models that accept images and conversational prompts together [4] shifted the economics. Questions such as "Which component is this?" or "Read this warning label aloud" no longer need a custom pipeline; they can be sent as zero-shot queries to a general cloud model.

6.2 Egocentric vision and the snapshot design

First-person video differs from ordinary vision data. It brings fast head motion, motion blur, hands that block the view and sudden changes in lighting. Large corpora such as Ego4D, with 3,670 hours of video from 931 wearers worldwide [13], exist because models trained on third-person footage generalize poorly to the wearer's viewpoint.

A snapshot-on-demand design sidesteps the tracking problem. The wearer holds their gaze steady and triggers a capture deliberately, so no processing of video over time is needed. The cost is the loss of continuous, proactive memory. A benchmark problem follows as well: existing egocentric benchmarks center on high-bitrate continuous video and say little about how accurate a single low-resolution still from a cheap microcontroller sensor can be.

6.3 Edge-cloud partitioning and reliability

Deciding what runs on the wearable, what runs on an edge relay and what runs in the cloud is the central distributed-systems question here, because it trades end-to-end delay against radio energy and device limits. Chen et al. measured latency across edge-offloading tiers for wearable cognitive assistants [14]. At the opposite extreme, DeepEye showed that several deep convolutional networks can run entirely on a compact wearable camera with a commodity mobile system-on-chip (SoC), at the cost of accuracy and considerable heat [15].

Ultra-low-power microcontrollers are far more constrained. Runtimes such as TensorFlow Lite Micro [16] handle small jobs like keyword spotting or simple gesture recognition but cannot perform open-vocabulary visual reasoning. Sending the reasoning to a cloud foundation model is therefore a structural necessity rather than a design preference.

We could find almost no measurements, however, of the end-to-end latency of sending a compressed JPEG from a low-power microcontroller through a smartphone relay to a remote multimodal API.

Reliability deserves equal attention. Generative models are known to hallucinate [17], and commercial multimodal models can misidentify small component details or misread text. We expect these errors to grow with the degraded images that cheap lenses produce. Generated answers should therefore remain advisory, especially in safety-critical industrial work.

7. HUMAN-CENTRIC INDUSTRY 5.0 PERSPECTIVE

Industrial settings raise the stakes of all this. The European Commission's Industry 5.0 framework places human-centricity, resilience and sustainability at the heart of industrial automation, so that technology empowers operators rather than replacing them [18]. A decoupled, low-cost modular sensing wearable aligns with these pillars in several ways.

Because the user starts each capture, the operator remains the main decision-maker. Graceful degradation keeps local sensor telemetry running when the cloud link fails, which supports resilience. Reusing TWS earbuds that workers already own supports circular-economy sustainability, although the footprint of manufacturing the extra lithium-polymer batteries calls for a proper life-cycle assessment.

Data governance requires visible recording indicators and negotiated workplace consent, under legal frameworks such as India's Digital Personal Data Protection (DPDP) Act of 2023 [19] and the European General Data Protection Regulation (GDPR).

These alignments remain theoretical. We found no study that tests low-cost modular smart eyewear against Industry 5.0 productivity measures such as assembly cycle time, error reduction or subjective cognitive load.

8. COMPARATIVE ANALYSIS OF REVIEWED STUDIES

8.1 Survey of representative studies

Table 3 lists representative peer-reviewed studies on cognitive assistance, on-device edge computing, bystander privacy, technology acceptance and egocentric perception.

Table -3: Literature survey of representative verified studies

Study	Method and platform	Key finding	Limitation
Ha et al. (2014) [12]	Multi-tier cognitive assistance on Google Glass with a nearby server tier	Tight latency bounds for compute-heavy assistance; degrades gracefully when tiers fail	Purpose-built engines per task; needs nearby servers; Glass-era hardware
Chen et al. (2017) [14]	Empirical latency measurement of edge-based wearable assistants	Quantifies how computation placement shapes end-to-end delay	Continuous guidance, not single-shot question answering
Mathur et al. (2017) [15]	On-device execution of multiple CNNs on the DeepEye wearable camera	Five deep vision models run entirely on the device	Accuracy declines; recognition only; heavier than a microcontroller node
Denning et al. (2014) [2]	In situ study with bystanders of AR-glasses recording	Bystanders want notice of recording and the ability to refuse it	Qualitative; early-generation devices
Steil et al. (2019) [9]	Scene and eye-movement features trigger a camera shutter; 17 participants	Preserves privacy in the tested setting	Needs eye tracker; small sample; prototype
Rauschnabel et al. (2018) [8]	Survey-based structural model on AR smart glasses	Privacy threat to others is a distinct adoption barrier	Stated intention, not observed behavior
Zuidhof et al. (2025) [10]	Structural equation modeling; health-care professionals	Social-interaction constructs improve explanation; fewer threats than expected	Anticipated adoption; clinical context only
Grauman et al. (2022) [13]	Large-scale data collection and benchmarks (3,670 h, 931 wearers)	Benchmark suite for egocentric perception	Continuous video; not low-cost snapshots

Read together, these studies fall into three silos that rarely speak to one another. Acceptance and privacy studies rely almost entirely on cross-sectional, self-reported surveys about hypothetical or outdated prototypes. They capture what people say they would do, not what they do over time, and they cannot show how attitudes change once the hardware becomes discreet. Systems studies measure latency, memory and scheduling carefully, but on specialized infrastructure or high-end mobile SoCs built for continuous streaming. Privacy engineering proposes ingenious mechanisms such as hardware shutters, then tests them on small groups in the laboratory.

8.2 Research trends

Three trends stand out. Hardware has moved from bulky optical display projection toward conversational, assistant-driven eyewear. Behavioral modeling has moved from user-centric usefulness frameworks toward socio-technical models that include bystanders. And systems design has moved from purely centralized cloud offloading toward multi-tier edge-cloud partitioning.

Some findings appear to contradict each other, such as the lower privacy concern among clinical staff [10] compared with the general public [8]. We read this as a difference in institutional norms rather than a conflict in the data.

9. RESEARCH GAPS AND FUTURE RESEARCH DIRECTIONS

9.1 Identified gaps

Six gaps emerge from the literature, spanning technical, methodological and regulatory ground (Table 4).

Technical systems gap. To our knowledge, no peer-reviewed paper systematically evaluates a decoupled modular architecture that links a low-cost microcontroller camera node (for example an ESP32-S3), a personal smartphone relay, existing TWS earbuds and a general-purpose cloud multimodal assistant as a single pipeline. Existing work examines the parts separately: edge offloading [12], [14], mobile-SoC vision [15] and TinyML inference [16]. No one has measured how they perform together. Recent mobile operating systems also terminate background processes aggressively and restrict Bluetooth Low Energy (BLE) permissions, which may create transport bottlenecks that the wearable literature has not examined.

Methodological gap. Acceptance research relies on cross-sectional surveys about hypothetical adoption [8], [10], so we do not know how stated intention compares with real use over time. In particular, no one has measured whether people keep asking visual questions as a daily habit once the novelty wears off. Research on bystander privacy is

also mostly qualitative and predates generative AI assistants [2].

Datasets gap. Large egocentric datasets such as Ego4D [13] are richly annotated, but they consist of high-resolution video from high-end cameras. The open literature offers no benchmark of single, deliberately framed low-resolution stills (equipment serial plates, component pinouts, low-contrast text) captured by budget microcontroller sensors and paired with spoken questions. Privacy datasets are likewise limited to small laboratory samples [9].

Scalability gap. No empirical scalability models exist for snapshot-based visual assistance. Cumulative server-side token costs and per-query inference energy are unquantified, as is the risk of depending on proprietary multimodal APIs whose model versions and latency can change without notice.

Security and privacy gap. Proposed mitigations such as physical privacy shutters [9] and recording notices [2] have not been validated on ultra-low-cost clip-on devices. The purpose-limitation and data-fiduciary duties in the DPDP Act 2023 [19] and the GDPR sit uneasily with cloud-based visual inference. It also remains unclear who is liable when a hallucinated answer [17] causes industrial damage or injury.

Performance characterization gap. End-to-end latency in a modular visual assistance pipeline divides into eight sequential stages: (1) trigger detection, (2) image capture, (3) JPEG encoding, (4) local radio transfer (BLE or Wi-Fi), (5) cellular cloud upload, (6) multimodal model inference, (7) text-to-speech (TTS) synthesis and (8) Bluetooth audio delivery. Early cognitive-assistance work aimed for sub-second responses to continuous video guidance [12], [14], whereas a conversational question can tolerate a longer wait. Even so, no one has published how the milliseconds are distributed across these eight stages, what 2.4 GHz coexistence between BLE and Wi-Fi does to them, or how cold-start latency compares with warm-socket latency. Fig. 4 shows the pipeline.

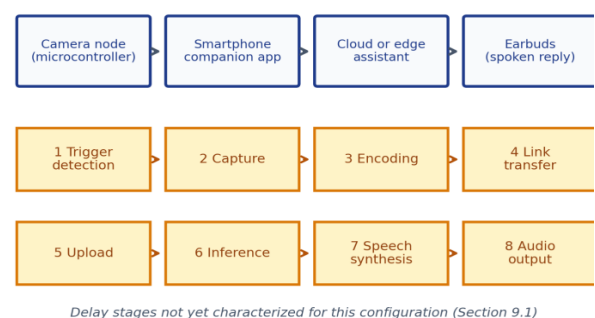


Fig -4: Modular visual-context pipeline (camera node, smartphone companion app, cloud or edge assistant, earbuds) and its eight delay stages, which remain uncharacterized for this configuration.

Table 4 sets the nearest existing evidence against the measurements that are missing.

Table -4: Research gaps identified in the reviewed literature

Gap area	Nearest existing evidence	What is missing	Basis
Technical	Edge offloading, on-device vision and microcontroller inference studied separately	Evaluation of camera node, smartphone, earbuds and assistant as one pipeline	[11], [12], [14], [15], [16]
Method	Self-reported intention studies; mostly qualitative bystander work	Behavioral and longitudinal evidence for assistant-equipped devices	[2], [8], [10]
Datasets	Ego4D continuous video; 17-participant privacy set	Labeled single low-resolution frames paired with spoken questions	[9], [13]
Scalability	Edge latency measured for continuous assistance only	Equivalent study for snapshot assistance; per-query energy and token cost	[14]
Security and privacy	Recording indicators and scene-aware shutters on prototypes	Validation on low-cost clip-on devices; liability and consent analysis	[2], [9], [17], [19]
Performance	Latency studies of continuous cognitive assistance	Stage-wise delay, radio coexistence, cold starts, accuracy on small text	[12], [14]

9.2 Future research directions

These gaps point to a five-part research roadmap, presented in priority order.

Vector 1: Willingness-to-pay (WTP) modeling. Econometric surveys should measure how much people would pay for a modular clip-on visual sensor, separating existing owners of TWS earbuds and compatible smartphones from non-owners. The whole idea rests on reusing hardware people already own, so price elasticity in emerging markets is the first thing to establish before any commercial case can be made.

Vector 2: Longitudinal acceptance studies that include bystanders. Multi-week field deployments with working clip-on prototypes should pair validated UTAUT2 measures [7] with objective measures of bystander comfort [8], [2]. The goal is to learn whether visual querying settles into a habit or fades once the novelty ends.

Vector 3: Joint latency-accuracy benchmarking. An open protocol should measure end-to-end latency and visual question-answering accuracy together, across links (BLE 5.x versus Wi-Fi), sensor resolutions (QVGA through UXGA), mobile OS background modes, and cold-start versus keep-alive sockets. Compression levels should be checked against optical character recognition (OCR) error rates to find the best balance between transmission speed and reasoning accuracy.

Vector 4: On-device TinyML privacy filtering and indicators. Microcontroller-class nodes should be tested running lightweight TinyML models [16] that clean images before transmission, for example by blurring them or by triggering a shutter when a face or monitor is detected. These should be paired with standardized, tamper-evident recording indicators.

Vector 5: Consented shop-floor Industry 5.0 pilots. Structured pilots should compare hands-free modular visual assistance with paper and tablet instructions in live production. Following the data-minimization principles of the DPDP Act 2023 [19], they should measure assembly completion time, component error rates and operator cognitive workload.

10. CONCLUSIONS

Smart eyewear needed more than a decade to travel from polarizing laboratory demonstrations to measurable mass-market traction. The literature suggests the delay had several overlapping causes. Early products asked wearers to adopt a conspicuous gadget and asked bystanders to accept a silent camera on someone's face that offered little a smartphone did not already do [2], [8]. The revival that began around 2023 arrived with normal-looking frames and conversational multimodal assistants that finally gave the head-worn camera daily value [4]. Growth is concentrated in a single corporate ecosystem [1], [5]. Because technical, ergonomic and aesthetic changes arrived together, the literature cannot separate their individual effects.

The literatures also fail to connect with one another. Acceptance research is limited to self-reported intention surveys. Systems research measures latency mostly for continuous workloads on specialized hardware. Privacy research validates its mechanisms only on small, artificial groups. The space between cheap audio-only frames and

premium camera glasses is exactly where these fields fail to meet.

So far, visual context at earbud-level cost has been demonstrated only by hobbyist maker kits [11], and no peer-reviewed study evaluates a fully modular commercial alternative. How long that gap lasts depends on component costs, mobile OS policies and manufacturer strategy, and it could close quickly if budget camera frames or camera-equipped earbuds spread.

Progress needs a sensible order of work: establish real willingness to pay and social comfort with visible capture, build shared latency and reasoning benchmarks, test on-device privacy features with bystanders, and run consented workplace pilots against human-centric Industry 5.0 criteria. Until then, modular visual sensing remains a well-motivated design hypothesis rather than a proven commercial solution.

REFERENCES

- [1] IDC, "Smart glasses surge: The XR market is rewriting its own rules," IDC Blog, Jun. 15, 2026. [Online]. Available: <https://www.idc.com/?p=115747> (accessed Oct. 15, 2026).
- [2] T. Denning, Z. Dehlawi, and T. Kohno, "In situ with bystanders of augmented reality glasses: Perspectives on recording and privacy-mediating technologies," in Proc. ACM CHI, 2014, pp. 2377–2386.
- [3] IDC, "India's smart wearable shipments fall 7.4% in Q2 2026 as premiumisation drives record 8.8% YoY ASP growth to US\$21.0," press release, Sep. 8, 2026. [Online]. Available: <https://www.idc.com/resource-center/press-releases/india-wearable-shipments-decline-q2-2026/> (accessed Oct. 15, 2026).
- [4] Gemini Team, Google, "Gemini: A family of highly capable multimodal models," arXiv:2312.11805, 2023.
- [5] IDC, "Smart glasses growth cools to 35% as Meta doubles down on distribution," IDC Blog, Sep. 16, 2026. [Online]. Available: <https://www.idc.com/resource-center/blog/smart-glasses-growth-cools-to-35-as-meta-doubles-down-on-distribution/> (accessed Oct. 15, 2026).
- [6] F. D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," MIS Quart., vol. 13, no. 3, pp. 319–340, 1989.
- [7] V. Venkatesh, J. Y. L. Thong, and X. Xu, "Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology," MIS Quart., vol. 36, no. 1, pp. 157–178, 2012.
- [8] P. A. Rauschnabel, J. He, and Y. K. Ro, "Antecedents to the adoption of augmented reality smart glasses: A closer look at privacy risks," J. Bus. Res., vol. 92, pp. 374–384, 2018.
- [9] J. Steil, M. Koelle, W. Heuten, S. Boll, and A. Bulling, "PrivacEye: Privacy-preserving head-mounted eye tracking using egocentric scene image and eye movement features," in Proc. ACM ETRA, 2019, pp. 1–10, doi: 10.1145/3314111.3319913.
- [10] N. Zuidhof, O. Peters, P.-P. Verbeek, and S. Ben Allouch, "Social acceptance of smart glasses in health care: Model evaluation study of anticipated adoption and social interaction," JMIR Form. Res., vol. 9, Art. no. e49610, Feb. 2025, doi: 10.2196/49610.
- [11] Based Hardware, "OpenGlass: Open source smart glasses," GitHub repository. [Online]. Available: <https://github.com/BasedHardware/OpenGlass> (accessed Oct. 15, 2026).
- [12] K. Ha, Z. Chen, W. Hu, W. Richter, P. Pillai, and M. Satyanarayanan, "Towards wearable cognitive assistance," in Proc. ACM MobiSys, 2014.
- [13] K. Grauman et al., "Ego4D: Around the world in 3,000 hours of egocentric video," in Proc. IEEE/CVF CVPR, 2022, pp. 18995–19012.
- [14] Z. Chen et al., "An empirical study of latency in an emerging class of edge computing applications for wearable cognitive assistance," in Proc. ACM/IEEE Symp. Edge Comput. (SEC), 2017.
- [15] A. Mathur et al., "DeepEye: Resource efficient local execution of multiple deep vision models using wearable commodity hardware," in Proc. ACM MobiSys, 2017, pp. 68–81.
- [16] R. David et al., "TensorFlow Lite Micro: Embedded machine learning for TinyML systems," in Proc. MLSys, 2021.
- [17] Z. Ji et al., "Survey of hallucination in natural language generation," ACM Comput. Surv., vol. 55, no. 12, Art. no. 248, 2023.
- [18] M. Breque, L. De Nul, and A. Petridis, Industry 5.0: Towards a Sustainable, Human-Centric and Resilient European Industry. Luxembourg: Publ. Office Eur. Union, 2021.
- [19] Government of India, "The Digital Personal Data Protection Act, 2023," Act No. 22 of 2023, Gazette of India, Aug. 2023.