Passthrough Interpretive Assistant: Revealing Hidden Intent and Bias in eXtended Reality with AI

Guhn Lee
Dept. of Industrial Design, KAIST
leeguhn@kaist.ac.kr

Anam Ahmad Khan
KAIST
anam.khan@kaist.ac.kr

Andrea Bianchi
Dept. of Industrial Design, KAIST
andrea@kaist.ac.kr

Abstract

Visual information density, amplified by generative AI, strains individuals' ability to assess message intent. We introduce the Passthrough Interpretive Assistant (PIA), an XR-based AI system that analyzes real-world visual media (e.g., posters, screens) and provides semantic overlays showing main points, intent, and potential bias. A pilot study found overlays slightly improved comprehension, raised awareness of persuasive elements in some cases, and required low cognitive effort. These results suggest that context-aware XR assistance can help users understand dense visual content, though further testing in real-world settings is needed.

Keyword

XR-based AI Integration, Augmented Reality, Bias Detection, Human–AI Interaction, Cognitive Support

1. Introduction

Everyday environments are filled with digital and physical visual media, often exceeding people's ability to evaluate what they see [1]. AI-generated content adds ambiguity, forcing individuals to judge accuracy, framing, and intent in real time [2, 3]. Algorithmic platforms such as social media amplify this challenge by tailoring content to users and increasing exposure to persuasive messaging [4].

Recent research has explored how XR and AI can help users process complex environments. Systems like AiGet [5] and Guided Reality [6] provide low-disruption, context-aware assistance via gaze, contextual sensing, multimodal interactions, and step-by-step guidance. Sensible Agent [7] and SituationAdapt [8] optimize overlay placement and adapt to attention. However, these systems focus on physical tasks or object understanding rather than helping users interpret the meaning or communicative intent of visual media.

To address this gap, we introduce the Passthrough Interpretive Assistant (PIA), an XR-based AI system that identifies posters and articles to provide spatially anchored overlays highlighting main points, intent, and potential bias. A pilot study explored how these overlays influence comprehension and interpretation. Results suggest that context-aware XR assistance may help users better understand dense or persuasive visual content while keeping cognitive load low.

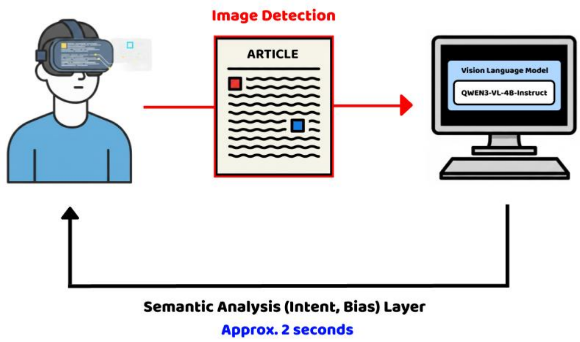

Fig 1. Passthrough Interpretive Assistant Workflow.

2. Related Work

Research informing the Passthrough Interpretive Assistant spans multimodal cognitive assistance, adaptive interfaces, and intent-aware agent modeling. Prior work establishes the feasibility of low-disruption XR support but focuses on physical tasks over semantic interpretation of visual media.

2.1 Contextual Interface Design in AR

Effective overlays must balance placement, load, and social acceptability. Sensible Agent demonstrates that proactive AR agents can adapt information and modality to reduce interaction effort [7], while SituationAdapt uses VLM/LLM (Vision Language / Large Language Model) reasoning to optimize placement and avoid occluding people or shared resources [8]. Managing user state is also critical: AttentionAR integrates EEG, IMU, and gaze to detect attention shifts and guide adaptive intervention [9]. Earlier work on multimodal and context-aware AR interaction further demonstrates how active assistance can be embedded into everyday tasks [10]. These efforts establish how AR systems can support users without disrupting task flow.

2.2 Proactive and Intent-Aware Modeling

Intent recognition and bias indication require richer modeling of how users interpret information. Satori offers a precedent by identifying user intent during physical tasks [11], while Contextually-Driven Prompts and agentAR show how AR agents can map natural-language rules to contextual actions and generate applications from high-level specifications [12, 13]. While these systems demonstrate flexible and proactive behavior, they remain focused on task execution and object interaction rather than interpreting communicative purpose, argument structure, or potential bias in visual media.

While prior XR systems anchor AI analyses to physical entities for scene understanding or task guidance, they focus on identifying and interacting with objects or physical processes. The challenge of helping users in extracting arguments, assessing communicative intent, and identifying bias remains yet to be explored. The PIA addresses this gap by combining high-resolution media capture, VLM-based media analysis, and spatially aligned overlays designed specifically for meaning-making within everyday environments.

3. Passthrough Interpretive Assistant (PIA)

3.1 System Overview

The PIA is an XR-based AI assistant that helps users interpret visual information in their surroundings. It combines the following:

1. **Passthrough AR capture & Real-time Vision Pipeline (Meta Quest 3)**: The headset provides the live feed and locally executes the computer vision pipeline (OpenCV) to detect and track rectangular surfaces (e.g., posters, screens).
2. **Remote Vision–language model (Qwen3-VL-4B-Instruct)**: Hosted on a high-performance local server, the model analyzes the captured media to generate summaries, classify communicative intent, and highlight bias. This architecture ensures the material is analyzed and returned to the user in a low-latency window, avoiding the unpredictable delays of cloud-based APIs.

The system identifies visual targets, transmits them for semantic analysis, and returns results as spatially anchored overlays. This allows users to read and interpret content in situ without losing awareness of the physical environment.

3.2 Interaction Flow

A typical interaction begins when the user puts on the headset. As they look around, the PIA identifies rectangular text or media surfaces and stabilizes them as world-anchored objects. The system then performs a high-resolution capture, sends it for vision-language analysis, and presents the result directly on top of the physical source material. The overlay updates as needed but remains spatially locked, preserving natural reading flow and minimizing visual disruption.

3.3 Implementation

The Passthrough Interpretive Assistant (PIA) uses a client-server architecture with a Meta Quest 3 headset acting as the client. The client application, developed in Unity with the OpenCV for Unity plugin, handles all real-time visual processing locally on the headset's passthrough feed. This pipeline performs edge detection, contour refinement, and geometric filtering to accurately detect and track rectangular surfaces (posters, screens). The client generates stable bounding boxes over these targets, maintaining their position via smoothing and persistence parameters to ensure stable world-anchored overlays.

Once a target is identified, the system captures a high-resolution crop of the region and sends it to a local Flask-based Python server running on macOS. The locally hosted Qwen3-VL-4B-Instruct performs OCR (optical character recognition) across 32 languages, extracts text, and handles multimodal vision–language reasoning. This local client-server setup completes the processing round-trip, from image capture to overlay delivery, within approximately 2 seconds. This rapid response is essential for maintaining a stable, responsive user experience without the latency overhead or network jitter of external requests.

Responses follow a strict color-coded classification scheme to indicate informational, transparent advertising, or manipulative content. The server returns structured JSON responses that are rendered as spatially anchored overlays on top of the original visual targets within the headset's passthrough feed. This integration enables real-time, context-aware semantic analysis while preserving the user's situational awareness and physical context.

4. Evaluation

Fig 2. Participant in Experimental Setup.

We conducted a within-subjects pilot study to explore how XR-based AI assistance influences comprehension and interpretation of real-world text. This pilot was intended to generate preliminary insights into how such systems can aid situational awareness, cognitive efficiency, and understanding of environments that are dense with visual information.

4.1 Participant Demographic

Participants (n = 8) ranged from 22 to 34 years of age, representing diverse educational levels (Bachelor's and Master's Degrees) and genders. Prior experience with AR/VR Technology was predominantly low, with seven participants self-identifying as "Beginner" users and only one participant identifying as "Moderate".

4.2 Procedure

Participants engaged with two articles on vaping/smoking and artificial sweeteners. These topics both fall within the food and health domain, and the articles were written with balanced, neutral framing that presented arguments on both sides. Each participant experienced both a reading-only condition and an AI-assisted condition in which spatially anchored overlays provided concise explanations and surfaced potential intent or bias cues. Article order and condition assignment were counterbalanced to reduce order effects.

The study began with a pre-experiment survey capturing demographics, prior knowledge, and initial opinions on both topics. During the reading phase, participants completed each article and answered items assessing their perceived comprehension, clarity, and confidence in interpretation. In the AI-assisted condition, additional questions captured how the overlays influenced understanding, trust, and interpretation. Participants also provided brief evaluations of each article's objectivity and trustworthiness. After both articles, a post-experiment survey measured any shifts in topic opinions and collected feedback on the XR-AI system's usability and cognitive demand.

Across these stages, the study captured participants' perceived comprehension, interpretation confidence, perception of intent or bias, opinion change, and subjective evaluation of the AI assistance. The within-subjects design allowed each participant to serve as their own control while counterbalancing reduced potential confounds.

5. Results

Given the small sample size, the findings are descriptive and highlight preliminary trends rather than definitive effects.

In the AI-assisted condition, self-reported comprehension increased slightly (Mean $\Delta = +0.37$, SD = 0.53), while interpretation confidence remained largely unchanged (Mean $\Delta = -0.13$, SD = 1.28) on the LIkert scale.

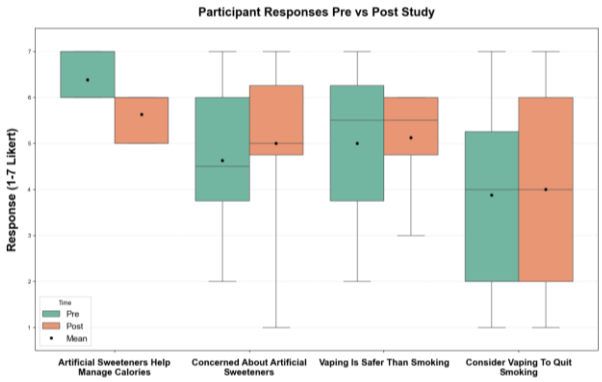

Fig 3. Participant Responses Pre vs. Post Study.

Participants' responses to "artificial sweeteners are a helpful tool for managing calorie intake" decreased on average by 1.25 points (SD = 0.52) after AI assistance. Responses concerning the safety of vaping showed minimal change (Mean $\Delta = -0.13$, SD = 1.13).

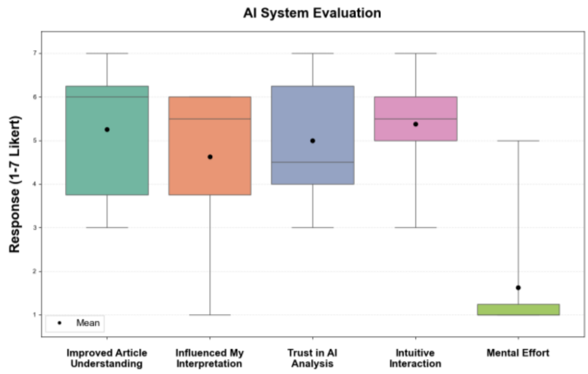

Fig 4. PIA System Evaluation Metrics.

Subjective usability metrics in the AI-Assisted condition reported the system as moderately helpful (mean 5.25, SD = 1.67) and trustworthy (mean 5.00, SD = 1.51). Participants found the system to be influential on their interpretation (mean 4.63, SD = 1.85), with four or more participants giving a rating of 6 or higher for the two metrics. Interacting with the XR system was generally intuitive (mean 5.38, SD = 1.19). The reported mental effort required to use the system was low (mean 1.63, SD = 1.40).

Notable individual cases illustrate the potential influence of AI assistance. Participant 3, who first used the system on the artificial sweeteners article, showed a drop in belief about the helpfulness of sweeteners, from 7 to 5, despite initially having low trust (3) in the source material. Participant 7 reported a decrease in self-confidence from 6 to 4 after AI assistance, coinciding with the system revealing "persuasive intent" in the article.

6. Discussion

The slight improvement in perceived understanding, combined with stable or reduced confidence, indicates that the system did not just simplify material. Instead, it introduced interpretive cues that encouraged users to reassess what they were reading.

This supports our initial premise that semantic overlays can reveal persuasive framing that may have gone unnoticed during rapid information processing. The system acted not as a correctness mechanism, but as support for interpreting communicative strategies and potential bias.

Individual cases illustrate both promise and risk. Some participants revised their stance or confidence after viewing neutral articles, suggesting that real-time interpretive cues can encourage calibrated skepticism. This represents a shift from earlier XR assistance systems such as AiGet and Guided Reality, which prioritize task clarity or object perception [5, 6]. The PIA instead operates at the qualitative level of meaning-making and communication analysis.

Compared with intent-aware systems like Satori [8] or adaptive frameworks such as Sensible Agent [9] and SituationAdapt [11], PIA extends context-aware support from predicting user actions to interpreting the purpose behind presented information. The attitude change observed in the artificial sweeteners article aligns with research showing that awareness of framing can shift judgments, underscoring the need for safeguards that prevent overreach or unintended influence.

Low reported mental effort suggests that semantic cues were integrated without disrupting cognitive flow. This complements prior work on unobtrusive and context-aware interfaces while bringing these ideas to higher-level content reasoning [5-8, 10]. The pilot highlights a methodological need to evaluate not only usability and comprehension, but also how semantic overlays affect interpretation in situ.

Overall, PIA demonstrates that XR-mediated interpretive support is feasible and impactful, yet sensitive to how cues are framed and delivered. Its influence on user interpretation raises questions around transparency, user agency, and appropriate boundaries for semantic guidance as such systems mature.

7. Limitations and Future Work

This work introduced the Passthrough Interpretive Assistant (PIA) as a proof-of-concept XR system for delivering on-demand, context-aware support when interpreting visual media. The prototype is limited by its reliance on a bulky headset, passthrough resolution, and dependence on device compute and tracking stability. These constraints reduce ecological validity and contributed to the study's controlled-lab setup.

While the pilot findings suggest that AI-generated cues can shape how users understand and evaluate content, the small sample size (n = 8), constrained environment, and participants' predominantly beginner-level experience with XR limit the strength of these conclusions. In particular, future studies with more experienced XR users are needed to determine whether the reported low mental effort reflects the system's design or a novelty effect associated with first-time use.

Interpretive overlays also raise ethical considerations, as they may unintentionally influence attitudes or encourage overreliance on model judgments. Safeguards such as clearer attribution, adjustable cue strength, and explicit acknowledgement of uncertainty are important as model capabilities grow. To strengthen the validity of these findings, subsequent studies should include a non-intelligent AR control condition, such as simple highlighting without AI analysis, to isolate the cognitive impact of AI inference from the visual effects of AR overlays. Future work should also leverage advances in lightweight XR hardware and efficient VLMs, and evaluate the system in naturalistic settings to better understand its technical, behavioral, and ethical impacts.

8. Conclusion

This study introduced the Passthrough Interpretive Assistant (PIA) as an early exploration of how XR systems and vision–language models might support users in interpreting visual media. The prototype showed that real-time overlays can shape users' understanding of persuasive content, but the findings are preliminary due to technical constraints and the controlled study environment. Rather than demonstrating effectiveness, the work identifies key opportunities and challenges for XR-based interpretive assistance.

The results point to the need for deeper inquiry into how such systems should be designed, evaluated, and governed. As hardware, computer vision, and interaction techniques advance, XR-mediated interpretive support may become more practical, yet its benefits and risks warrant careful empirical and ethical scrutiny. This proof-of-concept offers an early step toward that broader agenda by outlining a possible direction and surfacing the methodological and societal questions that future work will need to address.

Acknowledgement

This work was supported by the IITP (Institute of Information & Communications Technology Planning & Evaluation) - ITRC (Information Technology Research Center) grant funded by the Korea government (Ministry of Science and ICT) (IITP-2026-RS-2024-00436398).

Reference

1. Arnold, M., Goldschmitt, M. and Rigotti, T. Dealing with information overload: A comprehensive review. Frontiers in Psychology. 14. 1122200. 2023.

2. Gerlich, M. AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking. Societies. 15, 1. 6. 2025.

3. Balaskas, S., Konstantakopoulou, M., Yfantidou, I. and Komis, K. Algorithmic Burnout and Digital Well-Being: Modelling Young Adults' Resistance to Personalized Digital Persuasion. Societies. 15, 8. 232. 2025.

4. Metzler, H. and Garcia, D. Social Drivers and Algorithmic Mechanisms on Digital Media. Perspectives on Psychological Science. 19, 5. 735--748. 2024.

5. Cai, R., Janaka, N., Kim, H., Chen, Y., Zhao, S., Huang, Y. and Hsu, D. AiGet: Transforming Everyday Moments into Hidden Knowledge Discovery with AI Assistance on Smart Glasses. In Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems. ACM. pp. 1--26. 2025.

6. Zhao, A. Y., Gunturu, A., Do, E. Y.-L. and Suzuki, R. Guided Reality: Generating Visually-Enriched AR Task Guidance with LLMs and Vision Models. In The 38th Annual ACM Symposium on User Interface Software and Technology (UIST '25). ACM. 2025.

7. Lee, G., Xia, M., Numan, N., Qian, X., Li, D., Chen, Y., Kulshrestha, A., Chatterjee, I., Zhang, Y., Manocha, D., Kim, D. and Du, R. Sensible Agent: A Framework for Unobtrusive Interaction with Proactive AR Agents. In Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology (UIST '25). ACM. 2025.

8. Li, Z., Gebhardt, C., Inglin, Y., Steck, N., Streli, P. and Holz, C. SituationAdapt: Contextual UI Optimization in Mixed Reality with Situation Awareness via LLM Reasoning. In Proceedings of ACM UIST 2024. 2024.

9. Pei, Y., Huang, R., Zha, M., Wang, G., Wang, P., Kang, Q., Yang, Y. and Shen, H. T. AttentionAR: AR Adaptation and Warning for Real-World Safety via Attention Modeling and MLLM Reasoning. In Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology. 2025.

10. Brun, D. Multimodal and Context-Aware Interaction in Augmented Reality for Active Assistance. In Proceedings of the 20th ACM International Conference on Multimodal Interaction (ICMI '18). ACM, New York, NY, USA. pp. 506--510. 2018.

11. Li, C., Wu, G., Chan, G. Y., Turakhia, D. G., Quispe, S. C., Li, D., Welch, L., Silva, C. and Qian, J. Satori: Towards Proactive AR Assistant with Belief-Desire-Intention User Modeling. In CHI Conference on Human Factors in Computing Systems (CHI '25). ACM. 2025.

12. Dang, H., Lafreniere, B., Grossman, T., Todi, K. and Li, M. Authoring LLM-Based Assistance for Real-World Contexts and Tasks. In 30th International Conference on Intelligent User Interfaces (IUI '25). ACM. 2025.

13. Zhu, C., Hsia, S., Hu, X., Liu, Z., Shi, J. and Ramani, K. agentAR: Creating Augmented Reality Applications with Tool-Augmented LLM-based Autonomous Agents. In The 38th Annual ACM Symposium on User Interface Software and Technology (UIST '25). ACM. 2025.