Assistive AR System for Enhancing Human-Human and Human-Environment Interactions

Yunqi Guo
yunqiguo@link.cuhk.edu.hk
The Chinese University of Hong Kong
Hong Kong SAR, China

Abstract

Augmented Reality (AR) smart glasses promise transformative enhancements to human interactions and environmental awareness, yet face substantial challenges, including limited computational resources, constrained sensing capabilities, and inadequate intuitive interfaces. We introduce *Assistive AR*, a new category of AR systems designed explicitly for assistance to people's everyday lives, supporting both human-human and human-environment interactions through human-like sensing, efficient computation, and intuitive presentations. Specifically, *Sign-to-911* addresses communication barriers for deaf users during emergencies by providing real-time bidirectional Sign Language translation on the AR system. *SocialMind* enhances social interactions using proactive large language models (LLMs) to interpret and suggest responses based on conversational context. Further, *Sensor2Scene* and *Vivar* utilize generative models to visualize complex environmental sensor data to enhance the human-environment interaction. Moving forward, we will provide a general and scalable framework called *MASG-MCP* that enables users to create and customize tools on the Assistive AR system through a unified LLM agent.

CCS Concepts

• **Human-centered computing → Ubiquitous and mobile computing systems and tools**; **Accessibility systems and tools**.

Keywords

Assistive AR, Accessibility, Mobile Computing, Augmented Reality, Human-Computer Interaction, Large Language Models

ACM Reference Format:
Yunqi Guo. 2025. Assistive AR System for Enhancing Human-Human and Human-Environment Interactions. In *The 23rd Annual International Conference on Mobile Systems, Applications and Services (MobiSys' 25), June 23–27, 2025, Anaheim, CA, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3711875.3736666

1 Introduction

Humans engage daily in interactions with others and their environments. These interactions involve interpreting others' expressions, such as language or gestures, and environmental information like weather or air quality. However, these interactions are not always seamless. Individuals frequently encounter communication barriers due to hearing or speech impairments or social anxieties, causing delays in accessing critical services and complicating daily interactions. Moreover, interpreting environmental sensor data is challenging without specialized knowledge, especially when multiple analytical steps are involved. Current solutions fall short: traditional accessibility tools typically require cumbersome setups or lack real-time responsiveness, while sensor interfaces demand technical expertise, restricting their practicality.

Augmented Reality (AR) smart glasses hold transformative potential to overcome these barriers by supporting seamless, real-world interactions. Current AR hardware trends emphasize lightweight, wearable designs optimized for everyday use, exemplified by products like Meta Ray-Ban glasses featuring AI-driven photo capture, or waveguide displays by INMO and Rayneo targeting gaming and multimedia consumption. However, the true promise of AR glasses lies not merely in device-centric features but in facilitating real-world connections. First, AR glasses can acquire essential data on people's expressions, gestures, and environmental factors. Second, they can interpret this information in real-time to deliver contextually relevant insights, such as translating sign language or simplifying complex sensor data. Third, they intuitively present these insights, fostering seamless communication and environmental understanding. We define this emerging category of AR glasses designed to enhance daily interactions as *Assistive AR*.

Challenges. Despite their potential, AR glasses face significant hurdles. Current AR chips, such as Snapdragon AR2, offer substantially lower computational power than flagship mobile processors, restricting local processing for tasks like gesture recognition or real-time analytics. This limitation often forces reliance on cloud resources, introducing latency and privacy vulnerabilities due to data transmission. Sensing from a first-person perspective using cameras, microphones, and IMU (Inertial Measurement Unit) sensors poses additional challenges, as models trained on fixed-camera data struggle with dynamic inputs. Furthermore, without touchscreen interfaces, designing intuitive and proactive, hands-free interactions effective in diverse environments or user conditions becomes difficult. Finally, displaying complex information, such as sign language translations or multidimensional sensor data, requires immersive presentations beyond current smart-glass capabilities.

Contributions. To address these challenges, we adopt an interdisciplinary approach combining cross-device systems with specialized machine learning models. For accessibility, *Sign-to-911* [3] utilizes AI models optimized for linguistic features, enabling efficient two-way sign language translation on mobile devices. In social contexts, *SocialMind* [6] leverages proactive large language models (LLMs)—capable of anticipating conversational needs—and human-like sensing to enhance conversational interactions and engagement. For environmental interaction, *Sensor2Scene* [2] and

MobiSys' 25, Anaheim, CA, USA

ACM ISBN 979-8-4007-1453-5/25/06
https://doi.org/10.1145/3711875.3736666

Vivar [1] combine cross-modal embeddings, 3D Gaussian Splatting, and graphics shader-inspired Barycentric interpolation to generate immersive AR visualizations of sensor data. Finally, we propose leveraging a Mobile Assistive Smart Glass Model Context Protocol (*MASG-MCP*), which modularizes Assistive AR functions, streamlining efficiency and reusability across broader applications.

2 Assistive AR for Human-Human Interaction

AR glasses positioned in front of the eyes naturally facilitate human interactions without causing distractions. Leveraging this feature, we design Assistive AR systems to aid deaf-hearing communication and enhance social interactions.

Emergency services like 911 are often out of reach for deaf individuals due to communication barriers. Our solution Sign-to-911 [3] bridges this gap by using AR glasses for real-time, two-way translation between American Sign Language (ASL) and English. Rather than using heavy deep learning models, it employs compact AI/ML models specially optimized for ASL linguistic features, delivering both high accuracy and quick response times. Testing with six ASL signers using 911 call records showed that Sign-to-911 is a viable, scalable solution. The system works with off-the-shelf mobile devices and smart glasses, making it accessible without special infrastructure — filling the needs of emergency response systems.

In this system, we designed a customized infrastructure that uses AR glasses as both a sensory device to capture first-person-view signing and a display device for sign animations, despite their limited processing power. The main processing is offloaded to smartphones over Bluetooth bandwidth (150KB/s). This glass-mobile collaboration framework prioritizes efficiency and rapid response times, which informed the processing pipeline design of subsequent studies, including *SocialMind*.

Social interactions rely on subtle verbal and nonverbal cues that can be challenging to interpret. *SocialMind* [6] is an LLM-based AR system that provides proactive assistance during live conversations. Using first-person sensing (watching and hearing through AR glasses), it extracts cues like tone and body language and feeds them into LLM reasoning to generate real-time conversation suggestions. A user study demonstrated 38.3% higher engagement compared to baselines. *SocialMind*'s novelty lies in its in-situ, context-aware assistance, seamlessly enhancing social experiences without disrupting conversations. This work showcases AR's potential to improve interpersonal communication in daily life.

3 AR for Human-Environment Interaction

Assistive AR extends beyond human-human interaction, empowering users to perceive hidden aspects of their environment. Sensors are vital in measuring and interacting with the physical world [5]. As Internet of Things (IoT) sensors become more widespread, non-experts find it difficult to interpret their complex data. To solve this challenge, *Sensor2Scene* [2] and *Vivar* [1] create AR visualization systems that convert invisible sensor data into easy-to-understand 3D volumetric displays. These systems combine cross-modal embeddings, 3D Gaussian Splatting, and Barycentric interpolation—a graphics shader-inspired approach—to blend different types of sensor data into a unified embedding space. A validation study with 503 participants, including domain experts, demonstrated that the systems are effective and deliver consistent, faithful presentations.

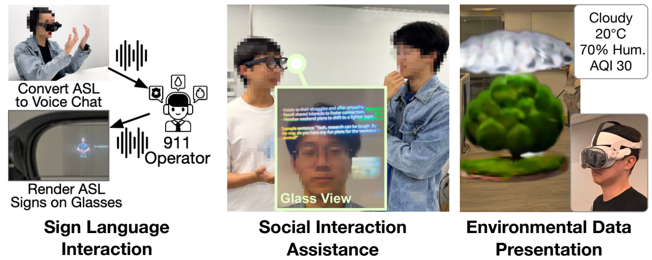

Figure 1: Assistive AR Systems

4 The Road Ahead

Future work will expand Assistive AR to serve a wider range of users, focusing on proactive assistance for the elderly, enhanced accessibility features for people with diverse needs, and intuitive applications for the general public. The Model Context Protocol (MCP) has shown its capabilities in assisting the interaction with sensor tools [4]. To unify fragmented AI tools on smart glasses and enable seamless, scalable assistance, we propose the Mobile Assistive Smart Glass empowered by MCP (*MASG-MCP*). This framework streamlines cross-device collaboration through LLM agents that coordinate sensing, machine learning, and feedback tools on personal devices, ensuring scalable task coordination. Through automated resource allocation, *MASG-MCP* will enable AR glasses to adapt dynamically to different user needs, delivering personalized support while addressing the interoperability on wearable systems.

5 Short Biography

Yunqi Guo received his Ph.D. from the University of California, Los Angeles, advised by Prof. Songwu Lu. He is currently a postdoctoral fellow at the Chinese University of Hong Kong, collaborating with Prof. Guoliang Xing. His research lies at the intersection of mobile and augmented reality systems, as well as visual-language interaction, emphasizing the development and deployment of innovative mobile systems that improve accessibility and enhance everyday human experiences.

References

[1] Yunqi Guo, Kaiyuan Hou, Heming Fu, Hongkai Chen, Zhenyu Yan, Guoliang Xing, and Xiaofan Jiang. 2024. Vivar: A generative ar system for intuitive multi-modal sensor data presentation. *arXiv preprint arXiv:2412.13509* (2024).

[2] Yunqi Guo, Kaiyuan Hou, Zhenyu Yan, Hongkai Chen, Guoliang Xing, and Xiaofan Jiang. 2024. Sensor2Scene: Foundation Model-Driven Interactive Realities. In *2024 IEEE International Workshop on Foundation Models for Cyber-Physical Systems & Internet of Things (FMSys)*. IEEE, 13–19.

[3] Yunqi Guo, Jinghao Zhao, Boyan Ding, Congkai Tan, Weichong Ling, Zhaowei Tan, Jennifer Miyaki, Hongzhe Du, and Songwu Lu. 2023. Sign-to-911: Emergency Call Service for Sign Language Users with Assistive AR Glasses. In *Proceedings of the 29th Annual International Conference on Mobile Computing and Networking*.

[4] Yunqi Guo, Guanyu Zhu, Kaiwei Liu, and Guoliang Xing. 2025. SensorMCP: A Model Context Protocol Server for Custom Sensor Tool Creation. In *The 23rd Annual International Conference on Mobile Systems, Applications and Services (MobiSys '25)*. ACM, Anaheim, CA, USA. doi:10.1145/3711875.3736687 3rd International Workshop on Networked AI Systems (NetAISys '25), June 27, 2025.

[5] Kaiwei Liu, Bufang Yang, Lilin Xu, Yunqi Guo, Guoliang Xing, Xian Shuai, Xiaozhe Ren, Xin Jiang, and Zhenyu Yan. 2025. TaskSense: A Translation-like Approach for Tasking Heterogeneous Sensor Systems with LLMs. In *Proceedings of the 23rd ACM Conference on Embedded Networked Sensor Systems*. 213–225.

[6] Bufang Yang, Yunqi Guo, Lilin Xu, Zhenyu Yan, Hongkai Chen, Guoliang Xing, and Xiaofan Jiang. 2025. Socialmind: Llm-based proactive ar social assistive system with human-like perception for in-situ live interactions. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 9, 1 (2025), 1–30.