Introduction to the First Workshop on Vision Foundation Models and Generative AI for Accessibility

Yapeng Tian
University of Texas at Dallas
Richardson, TX, USA
yapeng.tian@utdallas.edu

Yuhang Zhao
University of Wisconsin–Madison
Madison, WI, USA
yuhang.zhao@cs.wisc.edu

Jon E. Froehlich
University of Washington
Seattle, WA, USA
jonf@cs.uw.edu

Chu Li
University of Washington
Seattle, WA, USA
chuchuli@uw.edu

Yuheng Wu
University of Wisconsin–Madison
Madison, WI, USA
yuheng.wu@wisc.edu

Abstract

Vision foundation models and generative AI have recently made remarkable advances, enabling groundbreaking applications across computer vision, multimodal reasoning, and interactive AI systems. These models hold immense potential to enhance accessibility by powering assistive technologies that improve the daily lives of individuals with disabilities. At the same time, applying foundation models and generative AI in accessibility contexts presents key challenges, including fairness and bias, data scarcity, interpretability, and real-world deployment. This CV4A11y workshop brings together researchers from computer vision, human-computer interaction, and accessibility communities to identify challenges, explore emerging opportunities, and foster collaboration toward building inclusive, efficient, and responsible AI-powered accessibility systems.

1. Introduction

The rapid progress in vision foundation models and generative AI has reshaped the landscape of computer vision. Large vision and multimodal models (e.g., SAM [41], CLIP [68], GPT-4V [3], Gemini [86]) now demonstrate impressive generalization across tasks, from recognition and captioning to reasoning and generation. These advances open exciting possibilities for accessibility, such as real-time image description, automatic sign language translation, egocentric AI assistance, and intelligent mixed reality.

Despite these breakthroughs, accessibility remains insufficiently studied in computer vision research. Challenges such as bias against disability-related contexts, scarcity of representative datasets, and the gap between lab evaluations and real-world assistive use cases limit adoption. This workshop serves as a venue to articulate these challenges, highlight opportunities, and chart a roadmap for accessibility-centered computer vision and AI research. This year's workshop focuses on research intersecting vision foundation models, generative AI, and accessibility, including but not limited to:

- Vision foundation models for accessibility
- Image and video generation for accessibility
- Egocentric perception and wearable AI for accessibility
- Multimodal Large Language Models for accessibility
- Efficient foundation models for visual scene perception and understanding
- Automatic audio description for images and videos
- Sign language recognition, translation, and production
- AI-powered assistive technologies in mixed reality
- Human-centered evaluation and usability studies of AI-driven accessibility tools
- Fairness, ethics, and bias mitigation in AI for accessibility
- Challenges and opportunities in deploying AI-powered assistive systems in real-world settings

This workshop is highly relevant to the computer vision community, bridging cutting-edge research with real-world accessibility challenges. It will create new opportunities for impactful research, particularly for those working on scene understanding, image and video generation, egocentric perception, multimodal AI, and model efficiency, offering insights into how their expertise can contribute to more inclusive AI-driven systems. Additionally, it provides a unique platform for collaboration between computer vision researchers, accessibility experts, and industry developers. By fostering interdisciplinary discussions, it aims to drive breakthroughs in AI-powered accessibility and accelerate the development of real-world assistive technologies.

2. Related Workshops

The computer vision and accessibility research communities have a rich history of organizing focused venues to advance inclusive technologies. For example, the CVPR AVA workshops [2] have explored the intersection of accessibility, computer vision, and autonomous systems, with an emphasis on real-world assistive applications. The VizWiz Grand Challenge workshops [1] have provided a valuable platform for benchmarking and advancing technologies that support people with vision impairments, fostering progress in visual question answering and image accessibility. More recently, the Generative AI and Accessibility Workshop at CHI 2025 [28] reflects the growing interest in how generative models can support human–AI interaction and accessibility. We deeply value and build upon these past efforts. Each has played an important role in raising awareness, shaping research agendas, and mobilizing the community around accessibility challenges. However, accessibility-focused workshops within the core computer vision community (CVPR/ICCV/ECCV) remain limited, and few have directly addressed the opportunities and challenges presented by the latest advances in vision foundation models and generative AI. Our CV4A11y workshop aims to complement and extend these previous initiatives by creating a dedicated venue at ICCV 2025 to explore the technical aspects of integrating state-of-the-art foundation models with accessibility needs. Specifically, we focus on bridging the gap between rapid progress in computer vision and generative AI and their translation into scalable, robust, and inclusive accessibility solutions.

3. Workshop Organizers

The team has diverse expertise in computer vision, accessibility, and HCI with extensive experience in organizing conference workshops, ensuring they are well-equipped to run this multidisciplinary event.

Yapeng Tian is an Assistant Professor in the Department of Computer Science at the University of Texas at Dallas. His research focuses on solving core computer vision, computer audition, and machine learning problems and applying the developed learning approaches to broad AI applications, such as multimodal scene perception and generation [44, 58, 66, 67, 87], computational photography [88, 97, 98], accessibility [21, 47, 63], and healthcare [17, 53, 101]. Dr. Tian has been recognized with the 2024 Amazon Research Award and was selected for the AAAI 2023 New Faculty Highlights. His recent research has tackled a range of accessibility problems and developed new approaches to assist individuals with various disabilities, including audio description and assistive AR systems for blind and low-vision individuals [47, 63], multimodal LLM-based video analysis approaches for children with autism [17], and sign language production methods for those who are deaf or hard of hearing [20, 21]. Additionally, he has co-organized tutorials on Audio-Visual Scene Understanding at WACV 2021 and CVPR 2021, a workshop on Efficient Large Vision Models at CVPR 2024, and a workshop on Generative AI at NeurIPS 2024.

Yuhang Zhao is an Assistant Professor in the Computer Science Department at the University of Wisconsin-Madison. Her research focuses on augmented/virtual reality, accessibility, and intelligent interactive systems. She designs and builds AI-powered systems to recognize users' internal (e.g., gaze patterns for intents) [91, 92] and external (e.g., surrounding environment) [47, 108] contexts and generate suitable multi-modal feedback to empower people with disabilities in various daily tasks, such as reading [92, 107], visual search [108], cooking [47], navigation [112, 113], and social activities [110]. She also investigates and addresses potential issues brought by the emerging AI and extended reality (XR) technologies, such as privacy & security [79, 114], safety [103]), equity [102, 104], and accessibility [38, 109, 111]. She has rich experience hosting and organizing workshops for both academic researchers and local disability communities, including MR and Accessibility Workshop at ISMAR 2019, Open Access Tools and Libraries for Virtual Reality at IEEE VR 2023 & 2024, and AI-assisted Vision workshop for visual impairment communities.

Jon E. Froehlich is a Professor in the UW Allen School of Computer Science, a Sloan Fellow, NSF CAREER Awardee, and co-founder of Project Sidewalk [72]—an open-source project aimed at mapping and assessing every sidewalk in the world for accessibility using crowdsourcing and AI. He is a leading expert at the intersection of Human-AI systems and accessibility, including for sports participation [45, 46], sound awareness [23, 36], real-time captioning [30, 37], visual augmentations [47, 78], artwork description and interpretation [13, 14, 39] as well as indoor [81–83] and outdoor accessibility [51, 56, 72, 75, 96]. He was the general chair for ASSETS'22, the premier ACM conference for accessibility.

Chu Li is a third-year Ph.D. student in the UW Allen School of Computer Science. Her research explores the intersection of human-AI interaction and urban science, focusing on enhancing urban accessibility through interactive technology. Her recent publications include developing machine learning pipelines to improve data quality and crowdworkers' domain knowledge [50]; investigating perceptions of mobility barriers across disability groups and designing personalized maps [52]; exploring how digital civic tools can facilitate community accessibility advocacy [48]; and creating accessible geovisualization systems to enable broader participation in shaping inclusive urban infrastructure [49].

Yuheng Wu is a PhD student at the MadAbility Lab at the University of Wisconsin-Madison. His research interests include Human-Computer Interaction, AI-powered Interactive Systems, Augmented and Virtual Reality (AR/VR), and Accessibility. He is interested in building intelligent interactive systems that empower people. Currently, he is working on building AI-powered AR system that helps people with low vision to understand their surrounding environments and perform activities of daily living.

4. Brief Overview

Computer vision (CV) has become a key enabler of assistive technologies, transforming the way people with disabilities interact with the world. Over the past few decades, researchers have developed CV-based solutions across multiple domains of disability—ranging from helping blind individuals "see" their surroundings to interpreting sign language for the Deaf. To provide context for our workshop, we present a brief overview of representative efforts in computer vision for accessibility. We organize this discussion by disability category—visual, auditory, motor, and cognitive impairments—highlighting foundational works in each area. This overview is not intended as a comprehensive survey; rather, it highlights a selection of representative studies to stimulate discussion and illustrate the breadth of prior research. We acknowledge the many other exciting and impactful contributions not covered here.

4.1. Blindness and Low Vision

Early assistive technologies for blind people relied on nonvisual solutions such as Braille [24] and screen readers [19]. A major breakthrough was optical character recognition (OCR), which enabled printed text to be read aloud. Early OCR-based reading machines [74], and later document scanners [26], became essential tools that greatly expanded access to print. For people with low vision, magnification technologies such as CCTV readers [34] and digital screen magnifiers [8, 19] were critical aids, while contrast enhancement and font customization became standard accessibility features. Although computer vision was long envisioned as an "artificial eye," early vision-based aids faced technical and usability challenges and saw limited adoption.

The rise of smartphones and advances in deep learning marked a turning point. Mobile cameras enabled apps that could recognize objects, read signs, and describe scenes, while built-in accessibility features like magnifiers, high-contrast modes, and zoom functions improved usability for low-vision users. A milestone was VizWiz [7], which allowed blind users to snap photos and ask questions answered by remote humans—demonstrating strong demand and inspiring automated systems. By the mid-2010s, breakthroughs in object recognition [35, 43] and image captioning [90, 100], powered by large datasets and deep neural networks, enabled researchers to build smarter assistive vision models. The VizWiz Grand Challenge [31] further advanced research with a large dataset of blind-captured photos and questions, while commercial products such as Microsoft's Seeing AI [60] and OrCam MyEye translated academic progress into impactful tools.

Today, research emphasizes real-time, holistic scene understanding for blindness and enhanced visual accessibility for low vision. Modern systems [46, 47, 63, 94] integrate segmentation, recognition, captioning, and visual augmentation to provide richer and more flexible assistance. For example, CookAR [47] developed a head-mounted AR system with real-time affordance visualizations to support safe and efficient interactions with kitchen tools. Meanwhile, advances in multimodal foundation models such as CLIP [68] and GPT-4 Vision [3] highlight strong potential for multimodal assistance across blindness and low vision [32, 59].

Overall, assistive vision technology has progressed from single-task solutions like OCR or magnification to integrated, human-centered systems. Key challenges remain in delivering real-time assistance, personalizing support to individual needs, and building inclusive datasets that reflect the diversity of blind and low-vision users. The field is increasingly moving toward combining advanced computer vision with user-centered design to ensure outputs are practical and accessible in daily life—whether delivered through speech, sound, or visual augmentations.

4.2. Deaf and Hard-of-Hearing

For people who are Deaf or hard-of-hearing, early computer vision research primarily focused on bridging communication gaps through sign language recognition. Over the past decade, the field has advanced dramatically, driven by the availability of large-scale datasets and deep learning methods. Datasets such as PHOENIX14 [42], PHOENIX14-T [10], CSL-Daily [115], BOBSL [4], and How2Sign [18] have provided valuable resources for training models on continuous sign language recognition and translation. A representative milestone was the Sign Language Transformers [11], a transformer-based architecture that jointly recognizes sign sequences and translates them into spoken language, significantly improving translation performance over earlier approaches.

Today, the scope of research is broader and more sophisticated. Dedicated venues such as the SLRTP workshops [77] at ECCV and CVPR provide a platform for the computer vision community to share progress and foster collaboration. Beyond recognition, sign language production is also gaining attention, with avatar-based systems that generate signing from text to support two-way communication [21, 73, 80]. Computer vision is also enabling new forms of auditory accessibility. For instance, augmented reality captioning systems [54, 76] display live-transcribed speech in a user's field of view, enhanced with speaker localization or visual cues for sound sources.

The broader trend is toward multimodal fusion, combining video with audio or wearable sensor data for more robust accessibility. Looking forward, key challenges include capturing the full richness of natural sign languages, such as facial expressions and body language, achieving real-time performance on mobile and AR platforms, and ensuring inclusivity across diverse signing styles and linguistic variations. Now, computer vision for auditory accessibility has progressed from constrained prototypes to increasingly capable, multimodal systems with tangible real-world impact.

4.3. Motor Disabilities

Motor disabilities, including spinal injuries, neuromuscular conditions, and limb differences, can significantly affect mobility and motor control. Computer vision has opened new possibilities for hands-free control and environmental interaction for individuals with limited motor function. A pioneering example is Camera Mouse [6], which used a standard webcam to track head movements, enabling people with severe cerebral palsy or traumatic brain injury to control a mouse pointer. The system required no wearables, could adapt to any visible feature, and restored fundamental computer access. Today, similar webcam-based gaze and head-tracking tools are built directly into mainstream operating systems as accessibility options. Beyond pointer control, gesture recognition systems allow users to issue commands with simple actions such as blinking, winking, or raising eyebrows. In mobility support, computer vision has been applied to power wheelchair control, using cameras to track eye or head movements [57, 99]. Cameras mounted on wheelchairs can also enable autonomous navigation—detecting obstacles, sidewalks, curbs, and safe pathways [16]. These technologies enhance safety and independence by guiding users across both indoor and outdoor terrains. On the urban accessibility front, Project Sidewalk is a notable initiative that combines computer vision with crowdsourcing to assess sidewalk accessibility at scale [72]. The project collects imagery of sidewalks, crosswalks, and other urban infrastructure, then analyzes it to identify barriers. This creates maps and tools that help people—especially those using mobility aids like wheelchairs or canes—navigate the built environment safely, accounting for uneven surfaces, curb cuts, and other obstacles [52].

Across these applications, common themes include intuitive, non-invasive interaction, real-world relevance, and empowerment through autonomy. Yet challenges remain: ensuring real-time performance, adapting to individuals' unique abilities, and building datasets that reflect the diversity of bodies, mobility aids, and movement styles. From head-tracked pointer control to street-level navigation insights, computer vision is increasingly transforming independence and access for people with motor impairments.

4.4. Cognitive and Developmental Disabilities

Applying computer vision to cognitive and developmental disabilities is a relatively new and underexplored frontier, yet it has already demonstrated promising outcomes in both clinical care and daily support. A key line of research focuses on memory and task assistance for individuals with dementia or traumatic brain injury. For example, the COACH system [9, 61] pioneered vision-based prompting for people with Alzheimer's disease. Using an overhead camera, COACH monitored task performance—such as handwashing—and automatically detected when a user was confused or skipped a step. Clinical trials showed that COACH reduced caregiver intervention while helping participants successfully complete daily tasks. This vision-driven prompting paradigm exemplifies how computer vision can act as the "eyes" of an intelligent assistant, providing just-in-time support to foster independence.

Another active research direction addresses social cue interpretation for individuals with autism spectrum disorder. Mahcine learning approaches analyzing gaze patterns, such as reduced attention to faces, have shown effectiveness in distinguishing autistic from neurotypical children [55, 93]. However, many methods still rely on specialized eye-tracking hardware or controlled laboratory setups, limiting real-world deployment. Recent studies [17, 62, 69, 70, 95] are beginning to leverage computer vision to recognize autism-related behaviors, such as atypical gaze shifts, self-stimulatory movements, or repetitive behaviors, directly from naturalistic video recordings. These advances open the door to scalable, non-invasive tools for behavioral analysis and potential therapeutic support.

For individuals with intellectual disabilities or Attention-deficit/hyperactivity disorder (ADHD), researchers are beginning to explore "cognitive aids" that integrate computer vision with augmented reality [5, 105]. For instance, wearable AR glasses could recognize the environment and provide prompts for navigation or task completion, while smart home cameras could detect unsafe situations—such as a stove left on—and send alerts. Although these systems remain largely conceptual, related applications like fall detection and wandering alerts for dementia are already in practice. Common technical approaches include activity recognition to infer user intent, facial analysis for emotion and attention tracking, and personalization to adapt to individual needs. A notable trend is combining computer vision with context-aware computing: vision provides real-time situational awareness, while decision-making is guided by AI reasoning models or therapeutic input.

Ultimately, success in this domain relies on close collaboration between computer vision researchers and clinical experts. Progress is measured not only in algorithmic accuracy but also in human outcomes—safety, independence, and quality of life. Computer vision, applied thoughtfully, has the potential to serve as a powerful ally in supporting cognition, fostering autonomy, and improving care for people with cognitive and developmental disabilities.

5. CV4A11y Workshop Papers

The workshop invited both extended abstract and full paper submissions. Extended abstracts were limited to 2–4 pages and, while not published in the official conference proceedings, will be made publicly available on OpenReview. Full papers followed the ICCV style and were limited to 8 pages, excluding references, with accepted papers appearing in the ICCV workshop proceedings. Each submission was reviewed by at least two expert reviewers, and acceptance decisions were made based on their feedback. We are deeply grateful for the reviewers' thoughtful comments and timely efforts. In total, 15 papers were accepted, including 7 full papers and 8 extended abstracts.

The accepted works cover a broad range of topics at the intersection of computer vision, generative AI, and accessibility. These include assistive technologies for blind and low-vision users (e.g., audio description [89], real-time scene understanding [27], accessible image set description [71], and proactive visual interpretation [29]), sign language recognition and production [22, 65, 85], urban accessibility and mobility (e.g., navigation [12], curb ramp detection [64] and geospatial AI agents for accessibility queries [25]), accessible education [40], fairness and bias in disability-centered AI systems [33, 106], and multimodal large language models and generative applications [15, 84]. Collectively, these works highlight the growing momentum of this research community and set the stage for lively discussions on advancing vision foundation models and generative AI for accessibility.

6. Discussion and Remark

A growing trend in this field is cross-disciplinary collaboration. Leading vision conferences now include accessibility-related papers and workshops, while venues such as CHI and ASSETS emphasize user-centered design and real-world evaluation. Researchers today are not only advancing algorithms but also examining how people with disabilities actually use—and benefit from—these technologies in daily life. Approaches such as user studies, participatory design, and close engagement with disability communities are shaping the next generation of assistive AI systems to be practical, accepted, and empowering.

Computer vision for accessibility has progressed from niche prototypes into a vibrant field delivering life-changing tools. Early foundations in OCR and simple detectors have expanded into sophisticated AI assistants capable of perceiving and interpreting the world for those who need support. With ongoing advances in AI, sensing technologies, and inclusive design, the future holds great promise for vision-based solutions that further enhance independence, safety, and inclusion for people with visual, auditory, motor, and cognitive impairments.

This workshop, through invited talks, paper sessions, and panel discussions, seeks to attract strong participation from both academia and industry. Its overarching goal is to advance AI-driven accessibility solutions by leveraging vision foundation models and generative AI to design inclusive technologies that meaningfully benefit individuals with disabilities. By bringing together students, faculty, industry researchers, and accessibility experts from diverse backgrounds, the workshop will foster interdisciplinary collaboration, bridge gaps between AI and accessibility practice, and highlight real-world applications that improve quality of life. Ultimately, the aim is to help build a more inclusive society through cutting-edge research and innovation.

Acknowledgments

This work was supported by the National Eye Institute of the National Institutes of Health under Award Number R01EY037100. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

[1] Vizwiz grand challenge workshop. `https://vizwiz.org/workshops/`. 2

[2] AVA: Accessibility, vision, and autonomy meet (cvpr 2025 workshop). `https://accessibility-cv.github.io/`, 2025. 2

[3] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023. 1, 3

[4] Samuel Albanie, Gül Varol, Liliane Momeni, Hannah Bull, Triantafyllos Afouras, Himel Chowdhury, Neil Fox, Bencie Woll, Rob Cooper, Andrew McParland, et al. Bbc-oxford british sign language dataset. *arXiv preprint arXiv:2111.03635*, 2021. 3

[5] Ishit Arhatia, Bryan Abraham, and Godfrey Winster Sathianesan. Wandervision: Enhancing dementia care with artificial intelligence and augmented reality. In *2025 International Conference on Emerging Systems and Intelligent Computing (ESIC)*, pages 827–831. IEEE, 2025. 4

[6] Margrit Betke, James Gips, and Peter Fleming. The camera mouse: visual tracking of body features to provide computer access for people with severe disabilities. *IEEE Transactions on neural systems and Rehabilitation Engineering*, 10(1):1–10, 2002. 4

[7] Jeffrey P Bigham, Chandrika Jayant, Hanjie Ji, Greg Little, Andrew Miller, Robert C Miller, Robin Miller, Aubrey Tatarowicz, Brandyn White, Samual White, et al. Vizwiz: nearly real-time answers to visual questions. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology*, pages 333–342, 2010. 3

[8] Paul Blenkhorn, Gareth Evans, Alasdair King, S Hastuti Kurniawan, and Alistair Sutcliffe. Screen magnifiers: Evolution and evaluation. *IEEE Computer Graphics and Applications*, 23(5):54–61, 2003. 3

[9] Jennifer Boger, Jesse Hoey, Kate Fenton, Tammy Craig, and Alex Mihailidis. Using actors to develop technologies for older adults with dementia: A pilot study. 2010. 4

[10] Necati Cihan Camgoz, Simon Hadfield, Oscar Koller, Hermann Ney, and Richard Bowden. Neural sign language translation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7784–7793, 2018. 3

[11] Necati Cihan Camgoz, Oscar Koller, Simon Hadfield, and Richard Bowden. Sign language transformers: Joint end-to-end sign language recognition and translation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10023–10033, 2020. 3

[12] Alton Chao, Erika Maquiling, Edmund Chao, Roshan Sanjeev, Tonko Emil Westerhof Bossen, and Ross Greer. Automated context-aware navigation support for individuals with visual impairment using multimodal language models in urban environments. In *ICCV CV4A11y Workshop*, 2025. 5

[13] Arnavi Chheda-Kothary, Jacob Wobbrock, and Jon E. Froehlich. Engaging with children's artwork in mixed visual-ability families. In *The 26th International ACM SIGACCESS Conference on Computers and Accessibility*, 2024. 2

[14] Arnavi Chheda-Kothary, Ritesh Kanchi, Chris Sanders, Kevin Xiao, Aditya Sengupta, Melanie Kneisel, Jacob Wobbrock, and Jon E. Froehlich. Artinsight: Enabling ai-powered artwork engagement for mixed visual-ability families. In *Proceedings of the 30th International Conference on Intelligent User Interfaces (IUI '25)*, 2025. 2

[15] Ishant Yunay Chintapatla, Kazuma Choji, Naaisha Agarwal, Charles Duong, Vasu Sharma, Sean O'Brien, and Kevin Zhu. Corevqa: A crowd observation and reasoning entailment visual question answering benchmark. In *ICCV CV4A11y Workshop*, 2025. 5

[16] James M Coughlan and Huiying Shen. Terrain analysis for blind wheelchair users: computer vision algorithms for finding curbs and other negative obstacles. In *CVHI*, 2007. 4

[17] Shijian Deng, Erin E Kosloski, Siddhi Patel, Zeke A Barnett, Yiyang Nan, Alexander Kaplan, Sisira Aarukapalli, William T Doan, Matthew Wang, Harsh Singh, Pamela R. Rollins, and Yapeng Tian. Hear me, see me, understand me: Audio-visual autism behavior recognition. *IEEE Transactions on Multimedia*, 2024. 2, 4

[18] Amanda Duarte, Shruti Palaskar, Lucas Ventura, Deepti Ghadiyaram, Kenneth DeHaan, Florian Metze, Jordi Torres, and Xavier Giro-i Nieto. How2sign: a large-scale multimodal dataset for continuous american sign language. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2735–2744, 2021. 3

[19] Gareth Evans and Paul Blenkhorn. Screen readers and screen magnifiers. In *Assistive technology for visually impaired and blind people*, pages 449–495. Springer, 2008. 3

[20] Sen Fang, Chunyu Sui, Xuedong Zhang, and Yapeng Tian. Signdiff: Learning diffusion models for american sign language production. *arXiv e-prints*, pages arXiv–2308, 2023. 2

[21] Sen Fang, Chen Chen, Lei Wang, Ce Zheng, Chunyu Sui, and Yapeng Tian. Signllm: Sign language production large language models. *arXiv preprint arXiv:2405.10718*, 2024. 2, 4

[22] Sen Fang, Chen Chen, Lei Wang, Ce Zheng, Chunyu Sui, and Yapeng Tian. Signllm: Sign language production large language models. In *ICCV CV4A11y Workshop*, 2025. 5

[23] Leah Findlater, Bonnie Chinh, Dhruv Jain, Jon E. Froehlich, Raja Kushalnagar, and Angela Carey Lin. Deaf and hard-of-hearing individuals' preferences for wearable and mobile sound awareness technologies. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019. 2

[24] Emerson Foulke. Reading braille. *Tactual perception: A sourcebook*, 168, 1982. 3

[25] Jon E. Froehlich, Jared Hwang, Zeyu Wang, John S. O'Meara, Xia Su, William Huang, Yang Zhang, Alexander J. Fiannaca, Philip Nelson, and Shaun Kane. "does the cafe entrance look accessible? where is the door?" towards geospatial ai agents for visual inquiries. In *ICCV CV4A11y Workshop*, 2025. 5

[26] James Fruchterman. Accessing books and documents. In *Assistive technology for visually impaired and blind people*, pages 555–580. Springer, 2008. 3

[27] Loan Gia. Real-time scene understanding for blind users: Enhancing vision-language models for accessibility. In *ICCV CV4A11y Workshop*, 2025. 5

[28] Kate S Glazko, Mina Huh, Jazette Johnson, Amy Pavel, and Jennifer Mankoff. Generative ai and accessibility workshop: Surfacing opportunities and risks. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, pages 1–6, 2025. 2

[29] Ricardo Enrique Gonzalez Penuela, Felipe Arias-Russi, and Victor Capriles. Guiding multimodal large language models with blind and low vision visual questions for proactive visual interpretations. In *ICCV CV4A11y Workshop*, 2025. 5

[30] Greg Guo, Robin Yiru Yang, Johnson Kuang, Xue Bin, Dhruv Jain, Steven Goodman, Leah Findlater, and Jon E. Froehlich. Holosound: Combining speech and sound iden-

tification for deaf or hard of hearing users on a head-mounted display. In *Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)*, 2020. 2

[31] Danna Gurari, Qing Li, Abigale J Stangl, Anhong Guo, Chi Lin, Kristen Grauman, Jiebo Luo, and Jeffrey P Bigham. Vizwiz grand challenge: Answering visual questions from blind people. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3608–3617, 2018. 3

[32] Yu Hao, Fan Yang, Hao Huang, Shuaihang Yuan, Sundeep Rangan, John-Ross Rizzo, Yao Wang, and Yi Fang. A multi-modal foundation model to assist people with blindness and low vision in environmental interaction. *Journal of Imaging*, 10(5):103, 2024. 3

[33] Vishnu Hari, Kalpana Panda, Srikant Panda, Amit Agarwal, and Hitesh Laxmichand Patel. Who's asking? investigating bias through the lens of disability-framed queries in llms. In *ICCV CV4A11y Workshop*, 2025. 5

[34] Chandra M Harrison. Low-vision reading aids: reading as a pleasurable experience. *Personal and Ubiquitous Computing*, 8(3):213–220, 2004. 3

[35] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 3

[36] Dhruv Jain, Leah Findlater, Jamie H Gilkeson, Benjamin Holland, Ramani Duraiswami, Dmitry Zotkin, Christian Vogler, and Jon E. Froehlich. Head-mounted display visualizations to support sound awareness for the deaf and hard of hearing. In *Proceedings of the 33rd international conference on Human factors in computing systems*, 2015. 2

[37] Dhruv Jain, Rachel Franz, Leah Findlater, Jackson Cannon, Raja Kushalnagar, and Jon E. Froehlich. Towards accessible conversations in a mobile context for people who are deaf and hard of hearing. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*, pages 81–92, 2018. 2

[38] Tiger F Ji, Brianna Cochran, and Yuhang Zhao. Vrbubble: Enhancing peripheral awareness of avatars for people with visual impairments in social virtual reality. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility*, pages 1–17, 2022. 2

[39] Lucy Jiang, Jon E. Froehlich, and Leah Findlater. Making urban art accessible: Current art access techniques, design considerations, and the role of ai. In *The 4th Annual Workshop on The Future of Urban Accessibility: The Role of AI*, 2024. 2

[40] Ruslan Khrulev. Towards accessible education: A benchmark for ai-powered assistive grading. In *ICCV CV4A11y Workshop*, 2025. 5

[41] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4015–4026, 2023. 1

[42] Oscar Koller, Jens Forster, and Hermann Ney. Continuous sign language recognition: Towards large vocabulary statistical recognition systems handling multiple signers. *Computer Vision and Image Understanding*, 141:108–125, 2015. 3

[43] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012. 3

[44] Saksham Singh Kushwaha and Yapeng Tian. Vintage: Joint video and text conditioning for holistic audio generation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2025. 2

[45] Jaewook Lee, Devesh Sarda, Eujean Lee, Amy S. Lee, Jun Wang, Adrian Rodriguez, and Jon E. Froehlich. Towards real-time computer vision and augmented reality to support low vision sports: A demonstration of artennis. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST 2023)*, 2023. 2

[46] Jaewook Lee, Yang Li, Dylan Bunarto, Eujean Lee, Jaden Wang, Adrian Rodriguez, Yuhang Zhao, Yapeng Tian, and Jon E. Froehlich. Towards ai-powered ar for enhancing sports playability for people with low vision: An exploration of arsports. In *The 4th Workshop on Inclusion, Diversity, Equity, Accessibility, Transparency and Ethics in XR (IDEATExR)*, 2024. 2, 3

[47] Jaewook Lee, Andrew D. Tjahjadi, Jiho Kim, Junpu Yu, Minji Park, Jon E. Froehlich, Yapeng Tian, and Yuhang Zhao. Cookar: Affordance augmentations in wearable ar to support kitchen tool interactions for people with low vision. In *Proceedings of the 2024 ACM Symposium on User Interface Software and Technology*, 2024. 2, 3

[48] Chu Li, Katrina Oi Yau Ma, Michael Saugstad, Kie Fujii, Molly Delaney, Yochai Eisenberg, Delphine Labbé, Judy L Shanley, Devon Snyder, Florian PP Thomas, et al. "i never realized sidewalks were a big deal": A case study of a community-driven sidewalk accessibility assessment using project sidewalk. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, pages 1–18, 2024. 3

[49] Chu Li, Rock Yuren Pang, Ather Sharif, Arnavi Chheda-Kothary, Jeffrey Heer, and Jon E Froehlich. Altgeoviz: Facilitating accessible geovisualization. In *2024 IEEE Visualization and Visual Analytics (VIS)*, pages 61–65. IEEE, 2024. 3

[50] Chu Li, Zhihan Zhang, Michael Saugstad, Esteban Safranchik, Chaitanyashareef Kulkarni, Xiaoyu Huang, Shwetak Patel, Vikram Iyer, Tim Althoff, and Jon E Froehlich. Labelaid: Just-in-time ai interventions for improving human labeling quality and domain knowledge in crowdsourcing systems. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, pages 1–21, 2024. 2

[51] Chu Li, Rock Pang, Delphine Labbé, Yochai Eisenberg, Maryam Hosseini, and Jon E. Froehlich. Accessibility for whom? perceptions of mobility barriers across disability groups and implications for designing personalized maps.

In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, 2025. 2

[52] Chu Li, Rock Yuren Pang, Delphine Labbé, Yochai Eisenberg, Maryam Hosseini, and Jon E Froehlich. Accessibility for whom? perceptions of sidewalk barriers across disability groups and implications for designing personalized maps. *arXiv preprint arXiv:2502.19888*, 2025. 3, 4

[53] Guangyuan Li, Jun Lv, Yapeng Tian, Qi Dou, Chengyan Wang, Chenliang Xu, and Jing Qin. Transformer-empowered multi-scale contextual matching and aggregation for multi-contrast mri super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20636–20645, 2022. 2

[54] Jingya Li. Real-time augmented reality visual-captions for deaf and hard-of-hearing children in classrooms. In *2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pages 641–642. IEEE, 2023. 4

[55] Jing Li, Zejin Chen, Yihao Zhong, Hak-Keung Lam, Junxia Han, Gaoxiang Ouyang, Xiaoli Li, and Honghai Liu. Appearance-based gaze estimation for asd diagnosis. *IEEE transactions on cybernetics*, 52(7):6504–6517, 2022. 4

[56] Alex Liu, Kevin Wu, Minchu Kulkarni, Mikey Saugstad, Peyton Rapo, Jeremy Freiburger, Maryam Hosseini, Chu Li, and Jon E. Froehlich. Towards fine-grained sidewalk accessibility assessment with deep learning: Initial benchmarks and an open dataset. In *Extended Abstract Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility*, 2024. 2

[57] Wenping Luo, Jianting Cao, Kousuke Ishikawa, and Dongying Ju. A human-computer control system based on intelligent recognition of eye movements and its application in wheelchair driving. *Multimodal Technologies and Interaction*, 5(9):50, 2021. 4

[58] Tanvir Mahmud, Yapeng Tian, and Diana Marculescu. T-vsl: Text-guided visual sound source localization in mixtures. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 26742–26751, 2024. 2

[59] Daniela Massiceti, Camilla Longden, Agnieszka Slowik, Samuel Wills, Martin Grayson, and Cecily Morrison. Explaining clip's performance disparities on data from blind/low vision users. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12172–12182, 2024. 3

[60] Microsoft. Seeing ai: Talking camera for the blind. https://www.seeingai.com/. 3

[61] Alex Mihailidis, Jennifer N Boger, Tammy Craig, and Jesse Hoey. The coach prompting system to assist older adults with dementia through handwashing: An efficacy study. *BMC geriatrics*, 8(1):28, 2008. 4

[62] Farhood Negin, Baris Ozyer, Saeid Agahian, Sibel Kacdioglu, and Gulsah Tumuklu Ozyer. Vision-assisted recognition of stereotype behaviors for early diagnosis of autism spectrum disorders. *Neurocomputing*, 446:145–155, 2021. 4

[63] Zheng Ning, Brianna L Wimer, Kaiwen Jiang, Keyi Chen, Jerrick Ban, Yapeng Tian, Yuhang Zhao, and Toby Jia-Jun Li. Spica: Interactive video content exploration through augmented audio descriptions for blind or low-vision viewers. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pages 1–18, 2024. 2, 3

[64] John S. O'Meara, Jared Hwang, Zeyu Wang, Mikey Saugstad, and Jon E. Froehlich. Rampnet: A two-stage pipeline for bootstrapping curb ramp detection in streetscape images from open government metadata. In *ICCV CV4A11y Workshop*, 2025. 5

[65] Katerina Papadimitriou, Panagiotis Filntisis, George Retsinas, Gerasimos Potamianos, and Petros Maragos. Seeing in 2d, thinking in 3d: 3d hand mesh-guided feature learning for continuous fingerspelling. In *ICCV CV4A11y Workshop*, 2025. 5

[66] Weiguo Pian, Shentong Mo, Yunhui Guo, and Yapeng Tian. Audio-visual class-incremental learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7799–7811, 2023. 2

[67] Weiguo Pian, Yiyang Nan, Shijian Deng, Shentong Mo, Yunhui Guo, and Yapeng Tian. Continual audio-visual sound separation. *Advances in Neural Information Processing Systems*, 37:76058–76079, 2024. 2

[68] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PmLR, 2021. 1, 3

[69] Shyam Rajagopalan, Abhinav Dhall, and Roland Goecke. Self-stimulatory behaviours in the wild for autism diagnosis. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 755–761, 2013. 4

[70] Guilherme Ocker Ribeiro, Mateus Grellert, and Jonata Tyska Carvalho. Stimming behavior dataset-unifying stereotype behavior dataset in the wild. In *2023 IEEE 36th International Symposium on Computer-Based Medical Systems (CBMS)*, pages 225–230. IEEE, 2023. 4

[71] Piera Riccio, Luis Domene García, and Nuria M Oliver. Towards accessible image set descriptions with imageset2text. In *ICCV CV4A11y Workshop*, 2025. 5

[72] Manaswi Saha, Mikey Saugstad, Teja Maddali, Aileen Zeng, Ryan Holland, Steven Bower, Aditya Dash, Sage Chen, Anthony Li, Kotaro Hara, and Jon E. Froehlich. Project sidewalk: A web-based crowdsourcing tool for collecting sidewalk accessibility data at scale. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019. 2, 4

[73] Ben Saunders, Necati Cihan Camgoz, and Richard Bowden. Progressive transformers for end-to-end sign language production. In *European Conference on Computer Vision*, pages 687–705. Springer, 2020. 4

[74] Lawrence A Scadden. Kurzweil reading machine: Evaluation of model one. *Journal of Visual Impairment & Blindness*, 72(10):415–418, 1978. 3

[75] Ather Sharif, Paari Gopal, Mikey Saugstad, Shiven Bhatt, Raymond Fok, Galen Weld, Kavi Dey, and Jon E. Froehlich. Experimental crowd+ai approaches to track accessibility features in sidewalk intersections over time. In

Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility, 2021. 2

[76] Deogratias Shidende, Thomas Kessel, Anna Treydte, and Sabine Moebs. A personalized captioning strategy for the deaf and hard-of-hearing users in an augmented reality environment. In *International Conference on Extended Reality*, pages 3–21. Springer, 2024. 4

[77] SLRTP Workshop Organizers. SLRTP: 3rd workshop on sign language recognition, translation and production (cvpr 2025). https://slrtpworkshop.github.io/, 2025. 3

[78] Lee Stearns, Leah Findlater, and Jon E. Froehlich. Design of an augmented reality magnification aid for low vision users. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)*, 2018. 2

[79] Sophie Stephenson, Bijeeta Pal, Stephen Fan, Earlence Fernandes, Yuhang Zhao, and Rahul Chatterjee. Sok: Authentication in augmented and virtual reality. In *2022 IEEE symposium on security and privacy (SP)*, pages 267–284. IEEE, 2022. 2

[80] Stephanie Stoll, Necati Cihan Camgoz, Simon Hadfield, and Richard Bowden. Text2sign: towards sign language production using neural machine translation and generative adversarial networks. *International Journal of Computer Vision*, 128(4):891–908, 2020. 4

[81] Xia Su, Ruiqi Chen, Richard Weiye Zhang, Jingwei Ma, and Jon E. Froehlich. A demo of diam: Drone-based indoor accessibility mapping. In *Extended Abstract Proceedings of the 2024 ACM Symposium on User Interface Software and Technology*, 2024. 2

[82] Xia Su, Daniel Campos Zamora, and Jon E. Froehlich. Rais: Towards a robotic mapping and assessment tool for indoor accessibility using commodity hardware. In *Extended Abstract Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility*, 2024.

[83] Xia Su, Han Zhang, Kaiming Cheng, Jaewook Lee, Qiaochu Liu, Wyatt Olson, and Jon E. Froehlich. Rassar: Room accessibility and safety scanning in augmented reality. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, New York, NY, USA, 2024. Association for Computing Machinery. 2

[84] Alexandre Symeonidis-Herzig, Ozge Mercanoglu Sincan, and Richard Bowden. Visualspeaker: Visually-guided 3d avatar lip synthesis. In *ICCV CV4A11y Workshop*, 2025. 5

[85] Md. Tariquzzaman, Md Farhan Ishmam, Saiyma Sittul Muna, Md Kamrul Hasan, and Hasan Mahmud. Prompting with sign parameters for low-resource sign language instruction generation. In *ICCV CV4A11y Workshop*, 2025. 5

[86] Gemini Team, Rohan Anil, Sebastian Borgeaud, Jean-Baptiste Alayrac, Jiahui Yu, Radu Soricut, Johan Schalkwyk, Andrew M Dai, Anja Hauth, Katie Millican, et al. Gemini: a family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805*, 2023. 1

[87] Yapeng Tian, Jing Shi, Bochen Li, Zhiyao Duan, and Chenliang Xu. Audio-visual event localization in unconstrained videos. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018. 2

[88] Yapeng Tian, Yulun Zhang, Yun Fu, and Chenliang Xu. Tdan: Temporally-deformable alignment network for video super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3360–3369, 2020. 2

[89] Subhashini Venugopalan, Amy Pavel, Tyler Roper, Jimmy Tobin, Emily Wilson, Jenny Tan, Alicia Martin, and Anton Kast. Metrics for fine-grained evaluation of inline audio descriptions. In *ICCV CV4A11y Workshop*, 2025. 5

[90] Oriol Vinyals, Alexander Toshev, Samy Bengio, and Dumitru Erhan. Show and tell: Lessons learned from the 2015 mscoco image captioning challenge. *IEEE transactions on pattern analysis and machine intelligence*, 39(4):652–663, 2016. 3

[91] Ru Wang, Linxiu Zeng, Xinyong Zhang, Sanbrita Mondal, and Yuhang Zhao. Understanding how low vision people read using eye tracking. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–17, 2023. 2

[92] Ru Wang, Zach Potter, Yun Ho, Daniel Killough, Linxiu Zeng, Sanbrita Mondal, and Yuhang Zhao. Gazeprompt: Enhancing low vision people's reading experience with gaze-aware augmentations. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, pages 1–17, 2024. 2

[93] Shuo Wang, Ming Jiang, Xavier Morin Duchesne, Elizabeth A Laugeson, Daniel P Kennedy, Ralph Adolphs, and Qi Zhao. Atypical visual saliency in autism spectrum disorder quantified through model-based eye tracking. *Neuron*, 88(3):604–616, 2015. 4

[94] Yujia Wang, Wei Liang, Haikun Huang, Yongqi Zhang, Dingzeyu Li, and Lap-Fai Yu. Toward automatic audio description generation for accessible videos. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2021. 3

[95] Pengbo Wei, David Ahmedt-Aristizabal, Harshala Gammulle, Simon Denman, and Mohammad Ali Armin. Vision-based activity recognition in children with autism-related behaviors. *Heliyon*, 9(6), 2023. 4

[96] Galen Weld, Esther Jang, Anthony Li, Aileen Zeng, Kurtis Heimerl, and Jon E. Froehlich. Deep learning for automatically detecting sidewalk accessibility problems using streetscape imagery. In *Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)*, 2019. 2

[97] Bin Xia, Yulun Zhang, Shiyin Wang, Yitong Wang, Xinglong Wu, Yapeng Tian, Wenming Yang, and Luc Van Gool. Diffir: Efficient diffusion model for image restoration. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 13095–13105, 2023. 2

[98] Xiaoyu Xiang, Yapeng Tian, Yulun Zhang, Yun Fu, Jan P. Allebach, and Chenliang Xu. Zooming slow-mo: Fast and accurate one-stage space-time video super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 2

[99] Jun Xu, Zuning Huang, Liangyuan Liu, Xinghua Li, and Kai Wei. Eye-gaze controlled wheelchair based on deep learning. *Sensors*, 23(13):6239, 2023. 4

[100] Quanzeng You, Hailin Jin, Zhaowen Wang, Chen Fang, and Jiebo Luo. Image captioning with semantic attention. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4651–4659, 2016. 3

[101] Jiamiao Zhang, Yichen Chi, Jun Lyu, Wenming Yang, and Yapeng Tian. Dual arbitrary scale super-resolution for multi-contrast mri. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 282–292. Springer, 2023. 2

[102] Kexin Zhang, Elmira Deldari, Zhicong Lu, Yaxing Yao, and Yuhang Zhao. "it's just part of me:" understanding avatar diversity and self-presentation of people with disabilities in social virtual reality. In *Proceedings of the 24th international ACM SIGACCESS conference on computers and accessibility*, pages 1–16, 2022. 2

[103] Kexin Zhang, Elmira Deldari, Yaxing Yao, and Yuhang Zhao. A diary study in social virtual reality: Impact of avatars with disability signifiers on the social experiences of people with disabilities. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility*, pages 1–17, 2023. 2

[104] Kexin Zhang, Edward Glenn Scott Spencer, Abijith Manikandan, Andric Li, Ang Li, Yaxing Yao, and Yuhang Zhao. Inclusive avatar guidelines for people with disabilities: Supporting disability representation in social virtual reality. *arXiv preprint arXiv:2502.09811*, 2025. 2

[105] Shizhen Zhang, Shengxin Li, and Quan Li. Understood: Real-time communication support for adults with adhd using mixed reality. *arXiv preprint arXiv:2507.18151*, 2025. 4

[106] Xiantao Zhang. The escalator problem: Identifying implicit motion blindness in ai for accessibility. In *ICCV CV4A11y Workshop*, 2025. 5

[107] Yuhang Zhao, Sarit Szpiro, and Shiri Azenkot. Foresee: A customizable head-mounted vision enhancement system for people with low vision. In *Proceedings of the 17th international ACM SIGACCESS conference on computers & accessibility*, pages 239–249, 2015. 2

[108] Yuhang Zhao, Sarit Szpiro, Jonathan Knighten, and Shiri Azenkot. Cuesee: exploring visual cues for people with low vision to facilitate a visual search task. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 73–84, 2016. 2

[109] Yuhang Zhao, Cynthia L Bennett, Hrvoje Benko, Edward Cutrell, Christian Holz, Meredith Ringel Morris, and Mike Sinclair. Enabling people with visual impairments to navigate virtual reality with a haptic and auditory cane simulation. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, pages 1–14, 2018. 2

[110] Yuhang Zhao, Shaomei Wu, Lindsay Reynolds, and Shiri Azenkot. A face recognition application for people with visual impairments: Understanding use beyond the lab. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2018. 2

[111] Yuhang Zhao, Edward Cutrell, Christian Holz, Meredith Ringel Morris, Eyal Ofek, and Andrew D Wilson. Seeingvr: A set of tools to make virtual reality more accessible to people with low vision. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, page 111. ACM, 2019. 2

[112] Yuhang Zhao, Sarit Szpiro, Lei Shi, and Shiri Azenkot. Designing and evaluating a customizable head-mounted vision enhancement system for people with low vision. *ACM Transactions on Accessible Computing (TACCESS)*, 12(4): 1–46, 2019. 2

[113] Yuhang Zhao, Elizabeth Kupferstein, Hathaitorn Rojnirun, Leah Findlater, and Shiri Azenkot. The effectiveness of visual and audio wayfinding guidance on smartglasses for people with low vision. In *Proceedings of the 2020 CHI conference on human factors in computing systems*, pages 1–14, 2020. 2

[114] Yuhang Zhao, Yaxing Yao, Jiaru Fu, and Nihan Zhou. "if sighted people know, i should be able to know:" privacy perceptions of bystanders with visual impairments around camera-based technology. In *32nd USENIX Security Symposium (USENIX Security 23)*, pages 4661–4678, 2023. 2

[115] Hao Zhou, Wengang Zhou, Weizhen Qi, Junfu Pu, and Houqiang Li. Improving sign language translation with monolingual data by sign back-translation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1316–1325, 2021. 3