

LifeInsight: Design and Evaluation of an AI-Powered Assistive Wearable for Blind and Low Vision People Across Multiple Everyday Life Scenarios

Florian Mathis
University of St. Gallen
St. Gallen, Switzerland

University of Applied Sciences of the Grisons
Chur, Switzerland
florian.mathis@unisg.ch

Johannes Schöning
University of St. Gallen
St. Gallen, Switzerland
johannes.schoening@unisg.ch

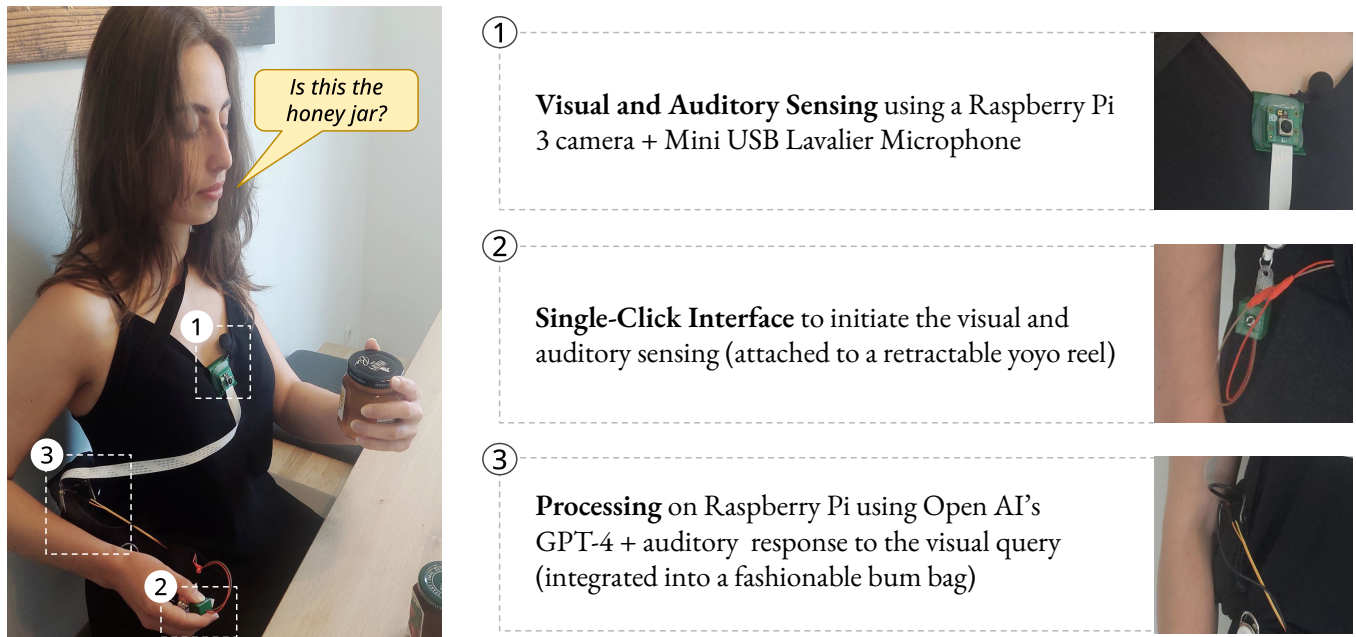


Figure 1: We present LifeInsight, an AI-powered assistive wearable equipped with an on-body camera, microphone, and a single-click interface for goal-oriented visual querying for blind and low vision (BLV) people. We evaluated LifeInsight across six scenarios, identified through multiple week-long video-based cultural probes with BLV people.

ABSTRACT

Assistive technologies (ATs) have the potential to empower blind and low vision (BLV) people. Yet, they often remain underutilised due to their immobility and limited applicability across scenarios. This paper presents LifeInsight, an AI-powered assistive wearable for BLV people that uses a wearable camera, microphone and single-click interface for

goal-oriented visual querying. To inform the design of LifeInsight, we first collected a corpus of BLV people's daily experiences using video probes and interviews. Ten BLV people recorded their daily experiences over one week using GoPro cameras, providing empirical insights. Based on these, we report on LifeInsight and its evaluation with 13 BLV people across six scenarios. LifeInsight effectively responded to visual queries, such as distinguishing between jars or identifying the status of a candle. Drawing on our work, we conclude with key lessons and practical recommendations to guide future research and advance the development and evaluation of AI-powered assistive wearables.

CCS CONCEPTS

- Human-centered computing → Empirical studies in accessibility; Accessibility technologies.

CHI '25, April 26-May 1, 2025, Yokohama, Japan

© 2025 Copyright held by the owner/author(s).

This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in *CHI Conference on Human Factors in Computing Systems (CHI '25)*, April 26-May 1, 2025, Yokohama, Japan, <https://doi.org/10.1145/3706598.3713486>.

KEYWORDS

Assistive Technologies, Wearables, BLV People

1 INTRODUCTION

Blind and low vision (BLV) people face numerous challenges in their everyday lives [32, 33, 52, 73, 76, 79, 107, 114, 117, 120, 129]. For example, when going grocery shopping, a common challenge for BLV people is to find specific products or differentiate between products [73, 76]. Other challenges include selecting clothing of a particular colour or identifying objects in the real world, including medical products or currency [32]. Although various forms of assistance are available to assist in such scenarios (e.g., human assistance or braille labels on the packaging of medication), current assistance, especially assistive technologies (ATs), are typically limited to specific scenarios. Taking widespread and frequently used talking colour detectors [88, 95] as an example, they only detect colours and the presence of light. Consequently, BLV people often use multiple ATs for assistance in their everyday lives. With the use of artificial intelligence (AI) and advances in smartphone apps like Seeing AI [7] and Be My Eyes [6], which utilise AI to describe people, text, currency, colour, and objects, the landscape of assistance has evolved significantly. Pioneered by Bigham et al. [26, 27] in VizWiz, where remote workers provide answers to BLV people's visual questions, AI-powered systems allow to identify medication, scan product barcodes, and distinguish between banknotes using a single handheld device [7, 38, 52]. Moving a step further, the transition from handheld devices to wearables [28], where technology is used while being worn, holds promise as it reduces interaction friction by freeing up BLV people's hands, which are often already occupied holding the white cane or the leash of a guide dog.

To contribute to the promising direction of wearable ATs, this paper presents the design, development and evaluation of LifeInsight, an AI-powered assistive wearable¹ that uses a wearable camera, microphone and single-click interface for visual querying. LifeInsight allows BLV people to ask visual questions (e.g., “Is this the honey jar?”, Figure 1) to then receive goal-oriented information [124], i.e., responses on their specific questions that they provided through speech input. To inform LifeInsight's design and evaluation, we first contribute a corpus of daily scenarios experienced by ten BLV people using video-based cultural probes for the duration of one week. The resulting video recordings and interviews highlighted the participants' challenges in localising, organising, understanding and interacting with real-world entities. For example, they encountered challenges in determining when a hot teapot has reached its fill level, distinguishing between different food products, or noticing another person standing next to them. While participants reported owning various ATs that could assist them, we learned from them and

from the literature [78, 98] that existing ATs often remain unused due to technological immaturity and inconvenience. Consequently, the participants expressed a desire for future ATs that seamlessly integrate into their daily routines, rather than remaining idle in a drawer.

Informed by the video-based cultural probes, we then empirically evaluated LifeInsight in a user study involving 13 BLV people to shed light on BLV people's usage behaviours and preferences when interacting with an AI-powered assistive wearable across six common everyday life scenarios, such as distinguishing between food products. Participants emphasised LifeInsight's potential to replace their smartphone as the current most important AT as they will no longer be required to take their phone out of their pocket, hold it in one hand, and operate it with the other hand. LifeInsight was perceived as highly useful across multiple everyday life scenarios that were previously identified in the cultural probes, with a success rate of above 80%. We conclude by discussing the opportunities and challenges for AI-powered assistive wearables and provide practical recommendations to support researchers and practitioners when designing, implementing and evaluating future AI-powered assistive wearables with BLV people.

Contribution Statement: We first contribute a corpus of BLV people's daily scenarios using video-based cultural probes [30, 51]. The video recordings using the cultural probes and the subsequent interviews build the foundation of the paper to understand *what* the existing challenges are (Figure 3) and *how* we can best build and evaluate LifeInsight, an AI-powered assistive wearable for BLV people, across multiple scenarios (empirical contribution [123]). Second, based on the insights obtained using the cultural probes, we contribute the design, implementation and evaluation of LifeInsight (artifact and empirical contribution [123]). LifeInsight assists BLV people across multiple daily scenarios using a single wearable technology, which implementation we made publicly available at <https://github.com/FlorianMathis/LifeInsight> to contribute to open-sourced designs of ATs [36] and to encourage replication studies and future research. Finally, we contribute (a) three key lessons learned from the cultural probes and the process of designing, implementing, and evaluating LifeInsight, and (b) practical recommendations for designing AI-powered assistive wearables for BLV people. These insights are particularly relevant for the HCI and accessibility communities as well as for leading industry projects that have recently highlighted the benefits of assistive wearables in the everyday lives of BLV people (e.g., the partnerships between Be My Eyes, Meta, and Microsoft [1, 3]). In summary, this paper contributes knowledge to the following two research questions:

RQ₁: To what extent can an AI-powered assistive wearable address BLV people's needs and preferences in multiple everyday life scenarios?

RQ₂: How do BLV people interact with an AI-powered assistive wearable across multiple everyday life scenarios?

¹The term “AI-powered assistive wearable” refers to a wearable prototype that uses artificial intelligence - specifically OpenAI's GPT-4 Vision API [5] - and the combination of visual input and speech input to assist BLV people.

2 RELATED WORK

We review literature that examined the challenges faced by BLV people and the current landscape of ATs.

2.1 Challenges of Blind and Low Vision People

Research on BLV people covers a substantial number of accessibility research, contributing to nearly half of the published works on accessibility at ACM conferences, such as CHI and ASSETS (2010 - 2019, > 43% [84]). For example, Brady et al. [32] reported on their findings of a study with 5,329 blind users where they explored the visual questions that blind people would like to have answered. Their taxonomy of visual questions highlights the importance of identifying objects in blind people's real-world surroundings, reading text in the physical space, and understanding the visual or physical properties of a real-world object. Rodrigues et al. [102] identified the major concerns, expectations, challenges, barriers, and experiences BLV people face with smartphones. They found the usage of smartphones is a tedious and long task for BLV people, but that they value the benefits of using technologies and that they are resilient and willing to improve their ability to use such technologies. Related work by Rector et al. [101] explored the challenges BLV people face with exercise technologies and argued corresponding systems should rigorously decide when and how to deliver auditory or haptic feedback. There is additional research that explored how BLV people navigate surface level changes [129], engage with visual artwork [17, 82], access information on the web [85], authenticate [19, 23, 33, 67], and participate in cooking activities [75, 79, 120].

In summary, various research efforts have contributed an understanding of BLV people's challenges and their needs, often by conducting interviews, contextual inquiries in controlled settings, and surveys [42, 49, 116, 128, 129]. What is missing is groundwork that provides insights into the combination of BLV people's daily scenarios and how these can inform the design, implementation and evaluation of an AI-powered assistive wearable. Building upon the literature that applied contextual inquiries in controlled settings [129] and explored visual questions asked by BLV people [26, 32, 52], we use cultural probes [30, 51] within the activities of daily living [81, Figure 1] (Figure 6 in Appendix A), such as doing the laundry or interacting with digital interfaces, to offer comprehensive insights into the daily scenarios encountered by BLV people and to inform, for the first time, the design and empirical evaluation of LifeInsight.

2.2 Assistive Technologies for Blind and Low Vision People

A result of the growing interest in accessibility research is the increasing number of ATs aimed to empower BLV people in various scenarios, including object recognition and social interaction [49]. Besides the help of another person or the use of a guide dog, ATs have become a central element in BLV people's daily lives. These include braille keyboards and displays [94, 118], audio-tactile labels [34, 53], smart

glasses [115], and apps (e.g., Seeing AI [7] or various accessible map applications [47]). Below, we review ATs from the broader HCI and accessibility literature.

As one of the pioneering works, VizWiz [26] offers a versatile solution for answering visual questions. A BLV person can take a photo with their phone, such as one of a grocery store shelf, and then ask a visual question to receive a response from a pool of workers. Building on the concept of visual scene understanding, Guo et al. [57] developed an interactive screen reader for real-world interfaces. Gonzalez et al. [52] and Chang et al. [38] created an AI-powered scene description app for visual answering. Rahman et al. [100] explored how automated hand-based spatial guidance can bridge the gap between guidance and execution. Their system assists BLV people in tasks such as writing on printed forms or drilling holes for hanging paintings. With Find My Things, Wen et al. [121] proposed an app that allows BLV people to find their personal items using teachable AI [65]. Others, such as Ahmetovic et al. [13] and Jain et al. [63], explored sonification of navigation instructions [13] and using street cameras [63] for navigation assistance.

The HCI and accessibility literature has proposed many more ATs for BLV people, highlighting a trend of novel ATs that range from generic visual question-answering applications (e.g., [26, 52]) to more specialised devices (e.g., FingerReader [106], VizLens [57], ForeSee [131]). Recent advancements integrate sophisticated AI capabilities, broadening the scope and functionality of ATs (e.g., Seeing AI [7], Be My Eyes [6], or WorldScribe [38]). Our exploration of a novel AI-powered assistive wearable for BLV people advances the field of user-centred ATs by combining: i) the benefits of wearables, which are designed to seamlessly fit into daily life and provide timely, context-specific information [54, 87, 97]; with ii) multimodal large language models (e.g., OpenAI's GPT-4 Vision [5]) for goal-oriented information [124] on BLV people's visual queries.

3 VIDEO-BASED CULTURAL PROBES

Prior to developing and evaluating LifeInsight, we center BLV people into our research process by using cultural probes to explore their daily experiences and inform the AT design- and evaluation process. *Cultural Probes* offer valuable insights into individuals' activities, concerns, and values [30, 51]. Developed in the late 1990s [51], they have since become a widely accepted method for contextual research within the accessibility, HCI, and design communities [30, 86]. In our work, we use a video recording device (i.e., a one-click Go-Pro camera), along with various straps and mounts, as the cultural probe. Participants were asked to provide insights into their everyday lives and the related challenges using the provided probes. Therefore, the resulting empirical data are represented by video recordings that were self-recorded by the participants. We conducted complementary interviews after the collection of the video recordings. While some of the insights from the use of the cultural probes align with the work by Brady et al. [32], which used a different methodology



Figure 2: For the cultural probes, we provided participants with a GoPro Hero 9 camera, including various straps and mounts. Participants received instructions on how to use the equipment and were assured of ongoing assistance.

in 2013, our video recordings provide access to more recent data (collected over a decade later), and allow us to more accurately describe the design- and evaluation process of LifeInsight without making assumptions about BLV people’s needs and preferences [89].

3.1 Cultural Probes and Scenarios

We equipped participants with a one-click video recording device, a GoPro Hero 9 Black with QuikCapture² (Figure 2). We used the GoPro Hero 9 Black due to its simplicity (single button press to start/stop a recording), its audio feedback when starting/stopping a recording, and the wide-angle lens, which further supports BLV people when recording their daily activities. Participants were free to select their preferred combination of straps and mounts to ensure comfort in the recording and not interfere with their daily experiences. For example, providing wearables like chest mounts and head mounts to seamlessly record daily activities ensures minimal interference with existing technologies, such as smartphones, which play an important role for many BLV people [102]. The majority of the participants decided to record the videos using the chest mount or the head mount, with some of the participants also using the tripod.

Following an introductory session, encompassing an interview to gather demographics and an overview of the data collection procedure, participants were introduced to the video-recording device and the different mounts. We encouraged the participants to focus their video recordings on challenges and scenarios that frequently appear throughout the

day, such as when dressing or doing house work (Figure 6 in Appendix A).

3.1.1 Ethics and Scenarios. Our work received ethical approval at the *University of St. Gallen*. To protect participants’ privacy and anonymity, the raw video footage was only reviewed by the primary researcher. Participants were asked to respect the privacy of potential bystanders. They were asked not to video record sensitive activities, such as bathing or toileting, to avoid causing discomfort or invading privacy. Participants were free to start and stop the video recording at any point throughout the week. For example, they might have video recorded various everyday life challenges on day 1, day 3, day 4, and day 7. We encouraged the participants to record as many daily scenarios and challenges as feasible using the activities of daily living as examples, including dressing, eating, doing housework, and other daily activities (Figure 6 in Appendix A). This allowed us to receive insights into BLV people’s various short-dated daily scenarios spread across a week.

3.2 Concluding Session: Interview

After a 1-week data collection phase, we scheduled a concluding session for a closing interview and to create an anonymised copy of participants’ probes. The concluding session was conducted at a venue chosen for participants’ convenience, e.g., at their home. The main objectives were to prompt participants to describe the scenarios they had captured on video and express gratitude for their valuable participation in our research. Appendix B provides an overview of the core interview questions.

²QuikCapture allows a user to automatically record video with a single button press, see the original documentation: <https://community.gopro.com/s/article/What-is-QuikCapture-and-How-Does-it-Work>, last accessed 31 Jul. 2024

Table 1: We report the sample’s demographics, including responses on the Short Set on Functioning questionnaire (WG-SS [96]). ○○○○ means no difficulty; ●○○○ some difficulty; ●●○○ a lot of difficulty; ●●●○ cannot do it at all. For P5 we provide additional data of their blind partner, who provided consent to be involved in the analysis, as described in Section 3.3.

Participant	Gender	Age	Vision	Vision	Hearing	Mobility	Cognition	Self-Care	Communication
P1	man	44	Low Vision	●●○○	○○○○	●●○○	●●○○	○○○○	●●○○
P2	woman	49	Totally Blind	●●●●	○○○○	○○○○	○○○○	○○○○	○○○○
P3	man	38	Totally Blind	●●●●	○○○○	○○○○	○○○○	○○○○	○○○○
P4	woman	46	Totally Blind	●●●●	○○○○	○○○○	○○○○	●●○○	○○○○
P5 P5.a	man woman	67 64	Totally Blind	●●●●	○○○○	○○○○	○○○○	○○○○	○○○○
P6	man	26	Totally Blind	●●●●	○○○○	●●●●	○○○○	●●○○	○○○○
P7	woman	41	Low Vision	●●○○	○○○○	○○○○	○○○○	○○○○	○○○○
P8	man	41	Totally Blind	●●●●	○○○○	●●○○	○○○○	○○○○	○○○○
P9	woman	47	Totally Blind	●●●●	○○○○	●●○○	○○○○	○○○○	○○○○

3.3 Recruitment and Participants

Participants were recruited through local societies for BLV people and project partners, including the *Swiss National Association of and for the Blind (SNABLING)*, *Schweizerischer Blinden- und Sehbehindertenverband (SBV, Sektion Ostschweiz)*, and *Lebenshilfe Vorarlberg*. All participants were reimbursed according to the local reimbursement policy. We recruited a sample of nine BLV people. P5’s blind partner, with whom P5 lives together (P5.a, female, 64, totally blind), contributed additional insights into P5’s and their daily challenges. Therefore, a total of 10 BLV people contributed video recordings during the one-week data collection. For simplicity, all insights from both P5 and their blind partner (P5.a) are collectively referred to as “P5” in the remainder of the manuscript.

Overall, we had 5 male and 5 female participants, with an average age of 46.3 (SD = 11.98; range: 26 - 67). **Table 1** provides an overview. Participants reported high familiarity with technology, reporting an average of 4.11 (Mdn = 4, SD = 0.33) on a 5-point Likert scale, where 5 represents very high familiarity. They reported to live alone in a house or flat (3), with their wife and kids (1), in a shared flat (1), with their blind partner (2), at their parent’s house (1), in a flat with their daughter (1), and in a shared care home (1). Although the majority of the participants reported to live with their disability since birth, there were several cases where their eyesight degenerated over time: P3 became blind after an operation 14 years ago, P4 had a vision of 4% at the age of eight and became blind three years ago, P5 became blind at around the age of 30 ($\approx$ 35 years ago), P6 became blind at the age of 20 ($\approx$ 6 years ago), and P8 became blind eight years ago.

3.4 Data Analysis

3.4.1 Cultural Probes. Before reviewing the cultural probes, we defined two key topics when analysing the videos: (1) description of daily scenario and (2) challenge of daily scenario. Systematically analysing the videos helped us to focus on similar key elements across all probes. The videos were accompanied by an audio track where participants expressed

their thoughts or provided descriptions of the scenarios and challenges they filmed. We now provide an overview of the key topics.

Description of Daily Scenarios. We first aimed to describe the video-recorded everyday life scenarios that we observed in each probe. Describing the scenarios in the first place helped us to identify the related challenges. Describing the activities also complemented our understanding of the scenarios and the learnings from the interviews, where the participants walked us through their daily experiences.

Challenge of Daily Scenarios. Having understood the scenarios, our objective was to understand the accompanying challenges. As we have conducted the concluding interviews before reviewing the video probes, we could identify the challenges through a combination of the video material, the additional audio descriptions, and the notes we took during the interviews.

Additional Observations. We wrote down additional observations (i.e., Memos [39]) where we report reflective thoughts of our experiences when interviewing BLV people and reviewing the cultural probes.

To contribute towards consistency and protect the participants’ privacy, the initial reviewing of the raw video material was done by the primary researcher who was in close contact with the participants. Three researchers then employed affinity diagramming [43], a commonly used method in HCI research [83, 113] and accessibility research [79, 119, 130], to cluster and organise the daily scenarios and the experienced challenges by BLV people. The lead researcher first provided the context of the collected data and then set up an in-person clustering session with two additional HCI researchers using Miro as a digital whiteboard [4]. The clustering lasted for approximately 1.5 hours and was performed on the shared digital whiteboard. Towards the end of the clustering, we brainstormed possible labels for the clusters and spent an additional 1 hour to discuss the preliminary affinity structure, the labels, and the items that we disagreed on until we reached a consensus. This procedure facilitated our ability to identify and cluster the core challenges and scenarios of

BLV people in their everyday lives. Table 4 in Appendix D shows an overview of the video recordings' metadata for each participant.

3.4.2 Semi-Structured Interviews. We audio recorded the interviews and took notes of the participants' comments. When asked about their daily scenarios that contribute significant challenges to their living, we transcribed and translated their comments for further analysis. The same three researchers who conducted the affinity diagramming on the video recordings then employed another affinity diagramming [43] on the semi-structured interview data to cluster the responses, following the same procedure as outlined in Section 3.4. We first provided context of the collected data and then initiated the clustering session using Miro [4]. Towards the end of the clustering, we added possible labels for the clusters and discussed the affinity structure. We then iterated again over the labels and the affinity structure. We concluded by discussing any remaining open points until we reached a consensus. This method facilitated our ability to identify the characteristics BLV people desire in ATs and learn about their "ideal" AT, which we report in Section 4.2.

4 RESULTS: VIDEO-BASED CULTURAL PROBES

To inform LifeInsight's design (Section 5) and its evaluation (Section 6), we first present the results collected using the cultural probes.

4.1 Challenges of BLV People in Everyday Life

The use of the probes highlighted six key scenarios, which we synthesise into a taxonomy in Figure 3. Our taxonomy overlaps with some challenges of Brady et al. [32]'s taxonomy, such as identification, description and reading. Our objective is to provide additional insights into BLV people's lived experiences and use the challenge-related scenarios to inform the design and evaluation of LifeInsight in Section 5 and Section 6.

4.1.1 Spatial Localisation and Organisation. The video recordings highlighted various challenges of BLV people when localising and organising entities. For example, P9 faced difficulties in locating and picking up an object they dropped on the floor. Related, P3 faced challenges in localising the ball when playing with their daughter. P5's probes shed light on the challenges encountered while working with their partner in the kitchen, particularly instances where they accidentally collided with one another. Other examples included the localisation of the roses to cut (P2), finding the different ingredients for a meal on the kitchen counter (P6), or positioning and localising food on a plate or in a bowl (P1, P9).

Summary of Challenge 1

The video recordings highlighted that localising and organising entities, encompassing physical objects and other humans, represents a significant challenge for BLV people.

4.1.2 Understanding the Current State. When P4 prepared a bowl of cereal with banana and milk, they expressed difficulties in adding the right amount of milk and cereal. Related, P8 commented they face difficulties in understanding the right mix ratio, e.g., when combining chicken with pineapple and rice. Additional examples included a recording by P2, where they prepared an indoor fireplace and mentioned it is challenging for them to know when adding more firewood makes sense as they are unaware of the fire's current state. In another recording, P2 had to pause their kitchen blender several times when mixing ingredients as they were uncertain about the consistency. Other video recordings included determining if there is remaining food in a pan (P8), ascertaining when the teapot is filled with hot water (P9), evaluating the cleanliness of clothing (P3), and knowing when their cake (P4) or meat (P8) is properly baked/cooked.

Summary of Challenge 2

The video recordings demonstrated that the lack of visual feedback to understand an entity's state often introduces difficulties and uncertainties for BLV people.

4.1.3 Interacting with Devices and Interfaces. P8 shed light on their check-in routine at their workplace, facilitated by a touchscreen interface, requiring them to install tactile markers to enable their interaction with the device. While P3's video recordings showed they are accustomed to operating a particular elevator and know which buttons to press to reach the ground floor, they voiced to frequently encounter difficulties when using elevators equipped with touchscreens and visual representations of floor levels. Additionally, the recordings highlighted that many household devices, such as washing machines and ovens, introduce touchscreens that make the interaction for BLV people challenging. P3 mentioned it is difficult for them to accurately set the temperature of the oven. P4 and P8 added that due to the challenges they face with their washing machine, stove and oven, they usually abstain from using specific settings although they might be superior in some cases, e.g., when a quick washing cycle would be more eco-friendly. P2's video recordings raised concerns about the feedback modalities of technologies that are supposed to help them, explaining that while their robot vacuum cleaner emits various noises, they experience difficulties in accurately interpreting the sounds.

SPATIAL LOCALISATION AND ORGANISATION

Localizing and organizing entities, encompassing physical objects and other humans, introduces a significant challenge in the everyday life of BLV people.



Playing Ball Game: A person aims to localise the spatial positioning of a ball when playing a ball game in a park with their daughter.

Challenge: Localising the ball and the other person when engaging in a ball game.



Presenting Food on a Plate: A person wants to present food on a plate aesthetically for their guests.

Challenge: Organising the food on a plate when guests are over to visit.

INTERACTING WITH DEVICES AND INTERFACES

BLV people experience significant challenges when engaging and interacting with devices and interfaces, particularly when those lack clear feedback, e.g., in the form of physical buttons or precise audio instructions.



Checking in at Work: A person wants to check-in and check-out at their workplace using the system provided by the employer.

Challenge: Interacting with a touchscreen when checking in and out at work.

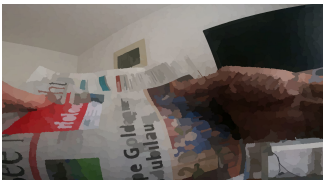


Interacting with the Oven: A person wants to set the temperature of the oven to a specific degree.

Challenge: Lack of interpretable tactile feedback when interacting with the oven.

COMPREHENDING WRITTEN INFORMATION

Despite significant technological advancements (e.g., OCR systems), assessing and comprehending information on physical objects is challenging for BLV people.



Reading Letters: A person wants to read a letter they received in their physical mailbox.

Challenge: Being able to understand from where the letter is and what the content of the letter is.



Understanding Instructions: A person wants to learn more about the washing that is best for a given clothing by reading the brand label.

Challenge: Lack of interpretable information on the brand label when aiming to understand the washing instructions of particular clothing.

UNDERSTANDING THE CURRENT STATE

The lack of feedback for BLV people to understand an entity's state often introduces challenges and uncertainties.



Preparing and Mixing Food: A person aims to cook chicken curry for their dinner. They combine chicken with pineapple and rice.

Challenge: Understanding the right mix ratio between different ingredients when preparing dinner.



Cleanliness of Clothing: A person wants to know if the clothes they wear are clean or dirty.

Challenge: Understanding the cleanliness of clothing.

COMPARING AND DIFFERENTIATING

Comparing, combining, and distinguishing between entities poses a challenge for BLV people, especially when dealing with everyday life objects that share common attributes, such as similar materials or shapes.



Sorting Socks: A person aims to compare different clothes to sort their socks.

Challenge: Differentiating between similar socks when doing the laundry.



Distinguishing Between Products: A person wants to distinguish different food products in their fridge.

Challenge: Distinguishing between food products that are stored in the fridge.

NAVIGATING IN (UN-)FAMILIAR ENVIRONMENTS

Navigating in environments is challenging for BLV people, especially in initially familiar surroundings where alterations disrupt navigational routines.



Walking in Lively Public Spaces: A person wants to walk from a bus stop to their workplace in a vivid environment.

Challenge: Construction sites or bicycles parked on the pavement to the way to the workplace introduce significant difficulties.



Moving Around at Home: A person wants to walk from one corner of their living room to the opposite corner of the same room.

Challenge: Changes in the living room, e.g., toys placed on the floor by the user's daughter, introduce significant mobility challenges.

Figure 3: A taxonomy of daily scenarios experienced by BLV people. The six core scenarios inform the evaluation of LifeInsight. A waterpixels filter was applied on the raw images to protect participants' identity and privacy.

Summary of Challenge 3

The video recordings indicated that BLV people experience significant challenges when engaging and interacting with devices and interfaces, particularly when those lack clear non-visual feedback.

4.1.4 Comparing and Differentiating. P5's recordings expressed challenges in filling the washing machine with clothing of the same colour. P8's recordings added the challenge of sorting socks and keeping track of two socks of the same type during the washing process. Others faced difficulties in differentiating between food products in the kitchen. For example,

P5 brought up the example of distinguishing between products in their fridge. P9 mentioned the example of grocery shopping and then arriving back at home, which introduces the challenge of differentiating between two similar products. Other video recordings showed challenges in distinguishing between spices and ingredients (P8, P4), or differentiating between playing pieces of a board game (P9).

Summary of Challenge 4

The video recordings highlighted that comparing and distinguishing between entities poses a significant challenge for BLV people, especially when dealing with everyday life objects that share common attributes, such as similar materials or shapes.

4.1.5 Comprehending Written Information. P3 and P5's recordings highlighted the challenges in reading physical letters and magazines, which require them to use time-consuming, and often inaccurate, optical character recognition (OCR) systems (e.g., OrCam [115]), or wait for human assistance. P5 explained the scenario where they need to wait for their son to visit before they can understand written information in a letter. P2's video recordings revealed that the challenge of comprehending information in written form is not only limited to reading magazines or letters, but extends to any type of information to be read on a physical object. For P2 and P5, the recordings highlighted challenges in comprehending washing instructions and product descriptions, including the washing instructions found on clothing labels.

Summary of Challenge 5

The video recordings emphasised that assessing and comprehending written information on real-world objects is challenging for BLV people.

4.1.6 Navigating in (Un-)Familiar Environments. P4 shared a recording where they walk from a bus station to work using a path they have rarely walked alone. Although no specific challenges were observed, P4 added it is challenging for them to adjust to adaptive environments. According to P4, construction sites or bicycles parked on the pavement are close-to-impossible for them to perceive. Related, P3's video recordings and their corresponding comments highlighted their familiarity with the living room setting, but that there are several occasions where their daughter's toys on the floor introduce significant challenges. P2 added they usually do not face any walking challenges at home, but that there are situations where both their hands are occupied and therefore they can no longer use their hands to guide them. They added navigation becomes even more challenging outdoors due to environmental noise, even in familiar settings such as in their garden.

Summary of Challenge 6

The recordings highlighted BLV people's challenges in navigating in environments, especially in initially familiar surroundings that become unfamiliar due to alterations or environmental noises.

4.2 Semi-Structured Interviews

When asked what characteristics an “ideal” future AT should manifest, the message of the participants was that future ATs need to be easy and fast to use, not be overloaded with too many features, and clearly improve their life. P9

articulated the imperative that any form of assistance must refrain from disconnecting them from reality, exemplified by not interfering with their hearing sense: “[a future AT] has to be very simple, it has to keep at least one ear free because I can't abstain from hearing” (P9). P2 and P3 emphasised the importance of future assistance being unobtrusive and omnipresent, seamlessly integrated into daily life without drawing attention or necessitating them to pick up dedicated devices.

“I shouldn't face difficulties when preparing to use the technology [...] there are pouring aids with an acoustic feedback, but I don't use them because it's kind of too annoying for me to get the device, put it in the glass, pour the water in the glass, and then clean the device” - P2

P8 shared their experience using OrCam MyEye, an AI-powered device that is mounted onto the arm of the glasses frame [115], emphasising the need for future ATs that provide greater comfort during extended use. Additionally, they voiced there exist some stigmatisation [11, 99, 110] and misunderstandings among sighted people, resulting in them feeling uncomfortable using such head-mounted devices in public.

“the stuff with the glasses [referencing to OrCam [115]] is quite good, but because the camera is on the side, it always pulls the glasses down on the right or left side. Also, in terms of comfort for wearing, it is not entirely pleasant, so it's not like you would say, ‘I'll go shopping with it.’ [...] And the next thing is, when you have a pair of glasses on, people get very confused. You have a white cane, symbolising that you can't see, but on the other hand, you're wearing glasses, and that's a bit... well, I tried it in the store with the products, and some people really came up to me and asked why I was wearing glasses... So there are misunderstandings. Maybe a different type of holder or something would be needed.” - P8

Expanding upon the essential traits that an “ideal” future AT should embody, we identified two primary directions highlighted by the participants:

Direction 1: Digital Vision Twin. P5 brought up the idea of a system that substitutes vision through real-time spatial descriptions of reality. Unlike a vibrating white cane, they envisioned a system that verbally identifies obstacles in their path, similar to how a guide dog would do it.

“it would be nice for mobility if there was a system that behaves like a guide dog [...] the dog never leads you into an obstacle, and if that is technically possible, then that would be a great help” - P5

Related, P4 brought up the example of a virtual guide dog, similar to NavCog3 [103] or RDog [37], that guides them through everyday life by describing their surroundings.

“a robot that walks ahead [...] and if stairs are approaching then, compared to a dog, the robot can tell me ‘stairs up, down’ [...]” - P4

P2 expressed interest in having access to a virtual agent that provides real-time information about their surroundings, and added they want to “replace their eyes” to allow them to go shopping and be able to explore discounts without the help of another person.

Direction 2: Human(-like) Assistance. There were comments on an “ideal” AT that represents, or acts like, a human. For example, P5 and P6 expressed their wish to have access to a human being, and that this is the best assistance they could get as it allows having continuous conversations.

“the most helpful thing, what is very difficult, is human assistance [...] if you can hire an assistant, that is of course always the greatest help. you can communicate with them” - P5

Additionally to the example of human assistance, P6 mentioned that a human-like assistant, such as in the form of a ChatGPT-extension³, might already be sufficient.

“[...] a system that simulates a human, but it’s difficult even with ChatGPT, you can use it to have something explained how it looks like but never so exactly; it’s still different to a human” - P6

In Section 5, we describe how the insights gained from the probes and the interviews inform the design and evaluation of LifeInsight, a first version of an AI-powered assistive wearable for BLV people. Through the probes and the interviews with BLV people, including the identification of the current challenges, needs, and wishes [89], we aim to shape future ATs that are useful for BLV people without creating irrelevant technical artefacts (i.e., “Disability Dongles” [62]).

5 LIFEINSIGHT: AN AI-POWERED ASSISTIVE WEARABLE FOR BLV PEOPLE

We first describe the concept and implementation of LifeInsight. We then contribute an empirical evaluation of LifeInsight across the six scenarios identified in the cultural probes (Figure 3 and Figure 5).

5.1 Concept and Implementation

With LifeInsight, we provide a single-click interface that allows BLV people to query targeted visual information across multiple real-world scenarios. LifeInsight serves as a human-like digital vision twin, offering visual information to BLV people; an aspect the probes and the interviews identified as important. Contrary to existing AI assistance, such as Be My Eyes [6], where BLV people take out their phone, navigate to the app, take a picture, and then provide written input, LifeInsight allows for targeted visual queries with a single-click interaction followed by speech input (Figure 1 and

Figure 4). LifeInsight’s strengths are its wearable, lightweight and flexible design, and its goal-oriented visual understanding by offering responses to BLV people’s specific objectives [124], which the probes and prior research deemed as essential. Williams et al. [122] found that blind people are “amenable to wearing such devices as long as they [are] not too large, heavy, or outlandish”. Furthermore, LifeInsight eliminates the need for additional interaction steps, such as taking out a smartphone or a dedicated AT, as it is worn on-body and consistently available for visual querying, aligning with the desire for “universal devices that enable seamless usage without the need for device switching” [49]. Early evidence in related fields highlights the potential of camera-based wearables, such as for memory aids for older adults [77] or to assist wheelchair users with shopping [71]. Recently, accessibility partnerships, like the one between Be My Eyes and Meta [1], suggest that AI-powered wearables will soon be available for millions of BLV people. While heads-on devices such as the OrCam [115] have found application, the comments by the participants (e.g., P8 in Section 4.2) and existing literature [29] suggest there is potential for the exploration of alternative form factors of future AI-powered assistive wearables.

In summary, LifeInsight can be worn as a wearable attached to clothing (e.g., clipped to the collar) or used as a handheld device, depending on BLV people’s preferences. While LifeInsight attached to clothing allows one-handed interaction, keeping the location of LifeInsight flexible can assist BLV people when pointing to specific objects they have visual questions about. This flexibility and interaction concept is derived from the use of Guiard’s kinematic chain model for human asymmetric bimanual cooperation [55, 56], where the user first sets the frame of reference and then initiates the visual querying pipeline using a single-click interface. In contrast to the Be My Eyes and Meta partnership [1], which allows BLV people to initiate the Be My Eyes’ “Call a Volunteer” experience entirely by voice commands on the Ray-Ban Meta Smart Glasses, our evaluation of LifeInsight contributes knowledge on the interaction concept and how an AI-powered assistive wearable is used by BLV people. While we designed LifeInsight as a lightweight and unobtrusive wearable that can be mounted onto clothing, its flexibility, including its form factor as well as its interactional affordances, still allows BLV people to mount LifeInsight on the frame of traditional glasses, aligning with the recent product developments in leading industry projects [1].

5.1.1 Implementation. LifeInsight encompasses three main components. Two of them, the camera + microphone for sensing, and the single-click interaction interface for initiating the visual querying process, are body-worn. The processing of the visual query takes place on a Raspberry Pi 4, providing a low-cost AI-powered assistive wearable for scientific explorations (Figure 4) that allows exploring the concept, benefits, and potential challenges of future AI-powered assistive wearables for BLV people.

³ChatGPT is a language model that can engage in natural language conversations and provide responses based on the input it receives, see <https://chat.openai.com/>, last accessed 31 Jul. 2024

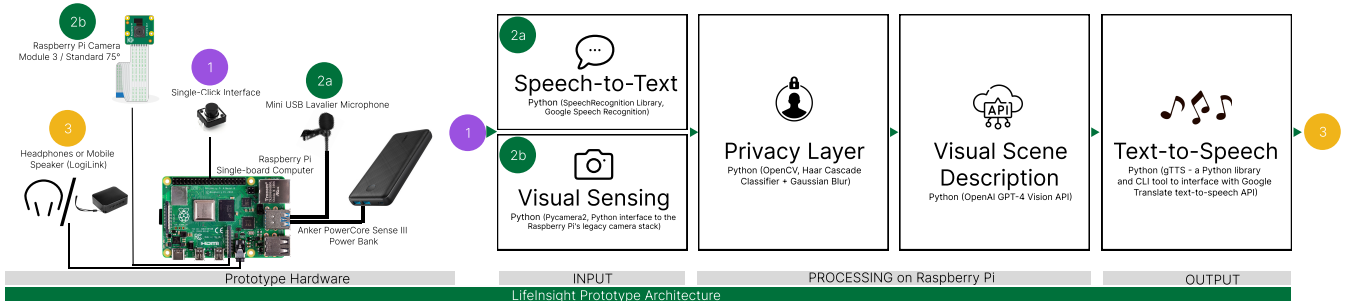


Figure 4: Implementation pipeline of LifeInsight, showing the individual hardware and software components.

Component 1: Visual and Auditory Sensing. For visual sensing of the real-world surroundings and to enable targeted visual querying, LifeInsight uses a Raspberry Pi 3 camera (Sony IMX708 image sensor) and a Mini USB Lavalier microphone. One reason LifeInsight offers speech input is that previous research has highlighted its speed compared to other input modalities, showing it to be nearly five times faster than using a keyboard with VoiceOver input [18]. All components that contribute to the processing of the real-world data, including a powerbank to power LifeInsight, are integrated into a fashionable bum bag (Figure 1).

Component 2: Single-Click Interaction Interface. The single-click interface for interacting with LifeInsight is mounted to a retractable yoyo reel that can be attached to any preferred body location. This flexibility has been made possible using a 3D printed case that provides BLV people a customisable interaction platform and allows us to better integrate aesthetics and customisation into the AT design [78]. A BLV person presses a button followed by speech input to initiate a targeted visual query. Auditory feedback informs about the button press and when LifeInsight is ready to receive input.

Component 3: Visual Querying Process. LifeInsight runs a Python script that processes the incoming data using a multimodal large language model, i.e., OpenAI’s GPT-4 Vision API [5], and then generates a response based on the visual question. For speech recognition, LifeInsight uses the SpeechRecognition library [9], which provides alphanumeric data that is then sent to the LLM together with a visual snapshot of the user’s real-world surroundings for contextual information. As system prompt for the multimodal LLM LifeInsight uses the following prompt on boot up: “Assisting a blind person in their everyday life. Responses should be brief and outline the requested information immediately within one sentence.”. The LLM output is then further processed using gTTS (Google Text-to-Speech) [44], a Python library to interface with the Google Translate text-to-speech API. All interactions are stored locally, including a visual snapshot of the real-world surroundings, the visual query, and the output by the LLM. Prior work by Akter et al. [15] has shown that BLV people are particularly concerned about the privacy of bystanders when capturing their surroundings. Therefore, for

privacy reasons, LifeInsight applies a haarcascade classifier and a Gaussian blur [41] to blur people’s faces (e.g., other bystanders) before further processing the visual data.

6 USER STUDY OF LIFEINSIGHT

Informed by the data collected using the cultural probes, we explore the extent to which LifeInsight can assist BLV people across multiple daily scenarios. In a within-subjects study with 13 BLV people, LifeInsight is evaluated across six common everyday life scenarios (Figure 5). The selection of the scenarios is further supported using literature that identified BLV people’s frequent daily challenges [26, 32, 58]. While some challenges faced by the BLV people may appear similar, such as reading information on devices/interfaces and comprehending written text, the video recordings and the interviews revealed these as distinct experiences with nuanced contextual differences, which we aimed to consider in the user study of LifeInsight and to contribute towards the breadth of the studied scenarios. While we present the individual findings for each scenario (e.g., in Section 7.1.1), we deliberately avoid direct comparisons between the scenarios. Instead, we aim to offer a comprehensive overview of LifeInsight’s strengths, weaknesses and its usage patterns among BLV people and across various everyday life scenarios. Figure 5 summarises the scenarios identified using the cultural probes, which we then used in LifeInsight’s evaluation.

- (1) **Localising Objects:** To study the use of LifeInsight for object localisation, we evaluate LifeInsight in the scenario of determining the presence of an object on the table (e.g., a watch). In this scenario, the participants were initially unaware of the watch’s presence and were tasked with determining whether a watch was on the table.
- (2) **Understanding the Current State of Objects:** To represent the scenario of understanding the current state of an object, we study LifeInsight in the scenario of determining the state of a candle. Participants were aware of a candle on the table and were tasked to identify its current status (i.e., lit or not). For safety reasons, the candle remained unlit throughout the study.

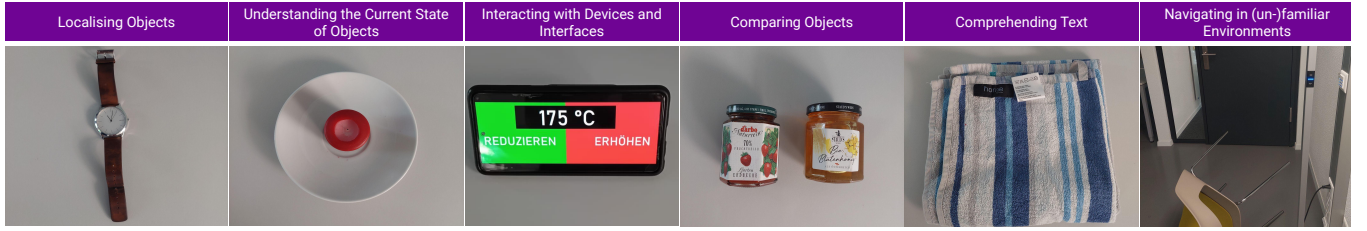


Figure 5: Overview of the scenarios that were identified in the cultural probes. Each object represents one of the scenarios (e.g., localising an object or comparing two objects). LifeInsight was evaluated in all scenarios.

- (3) **Interacting with Devices and Interfaces:** Despite growing interests in developing accessible digital interfaces (e.g., [68, 70]), reading information on displays remains challenging for BLV people. We investigate LifeInsight when aiming to identify the temperature written on a display (e.g., an oven). This scenario focused on the information displayed on the screen, without considering subsequent possible interactions, such as adjusting the temperature.
- (4) **Comparing Objects:** We evaluate LifeInsight in the context of distinguishing two food products who manifest similar form factors (i.e., a jam jar and a honey jar). The participants were instructed to determine which jar contained jam and which contained honey.
- (5) **Comprehending Text:** To assist BLV people in comprehending text in the real world, we study LifeInsight in the scenario of reading the washing information from the care label of a household linen (i.e., washing temperature of a towel). The participants were instructed to inform us about the washing temperature written on the care label of a towel.
- (6) **Navigating in (un-)familiar Environments:** Finally, we explore the use of LifeInsight to assist BLV people in recognising and identifying obstacles in front of a door (e.g., a chair and a cable). Compared to the task of localising a watch, in this scenario, the participants were aware of an obstacle in front of the door and were tasked to use LifeInsight to identify it.

Although there are numerous additional everyday life challenges that BLV people might face (e.g., when cooking [80, 120]), we focus on these six scenarios to align with the obtained insights in the cultural probes, reported in Section 4. These explorations offer important insights into BLV people’s behaviours and preferences when they use, for the first time, an AI-powered assistive wearable across multiple scenarios.

6.1 Recruitment and Participants

A new set of 13 participants was recruited through local societies and project partners, including some BLV people who had previously taken part in the cultural probes. The study received ethical approval at the *University of St. Gallen*. Participants were reimbursed according to the local reimbursement policy. We first conducted a pilot study with one

blind person and then recruited thirteen BLV people, including 7 male and 6 female individuals, with an average age of 51.77 (SD = 13.16; range: 27 - 67; see Table 2).

6.2 Data Collection and Analysis

The empirical evaluation of LifeInsight encompasses two phases: i) Scenario Exposure using LifeInsight; and ii) Semi-Structured Interviews. The order of the scenarios in Figure 5 was counter-balanced using a Latin Square [31]. Participants’ interactions when using LifeInsight were video- and audio recorded.

6.2.1 Phase 1: Scenario Exposure Using LifeInsight. We first introduced LifeInsight and explained all components, including the wearable camera, the microphone and the single-click interface for speech input. We then let the participants familiarise themselves with the system. Participants then experienced the six scenarios identified in the cultural probes (Figure 3 and Figure 5). For participants’ interaction behaviours, we first reviewed the recordings to identify their usage behaviours when interacting with LifeInsight. We then assessed participants’ queries with regard to each query’s proximity to the expected output. For example, when aiming to read the washing instructions of a towel, a BLV person might ask LifeInsight to output i) what it sees (*generic prompt*), ii) what it sees on the towel (*object-centred prompt*), or iii) the temperature they can wash the towel with (*targeted object-centred prompt*, i.e., goal-oriented information [124]). The classification adds to Brady et al. [32]’s taxonomy of visual questions and to Gonzalez et al. [52]’s understanding of BLV people’s goals when using AI-based apps. It allows us to understand *how* BLV people query for visual information when using LifeInsight for the first time. Furthermore, we asked participants to rate the wearable’s usability using the SUS [35], their perceived empathy of the wearable (PETS) [104], and their perceived workload using the raw NASA-TLX [59]. PETS encompasses two factors: emotional responsiveness (PETS-ER) and understanding & trust (PETS-UT) [104]. Additionally, we asked to rate the perceived satisfaction, trust and success when using LifeInsight on a 7-point Likert scale (1=low, 7=high). Finally, we report the appropriateness of LifeInsight’s descriptions given the visual and auditory sensing input. In line with Gonzalez et al. [52], we aimed to understand the appropriateness of LifeInsight’s responses given the information available in the photos processed by

Table 2: We report the sample’s demographics, including their experience with Seeing AI [7], as also done in the work by Gonzalez et al. [52], the assistance they had with them, and their affinity for technology (1 = low, 5 = high).

Participant	Gender	Age	Vision	Seeing AI Experience	Participant-Owned Assistance	Affinity for Technology
P1	man	37	Totally Blind	yes	White Cane, iPhone, Apple Watch	4
P2	woman	64	Low Vision	yes	White Cane, iPhone, Guide Dog	4
P3	man	67	Totally Blind	yes	White Cane, iPhone	4
P4	woman	65	Totally Blind	yes	White Cane, iPhone, Guide Dog, Milestone	4
P5	man	65	Totally Blind	yes	White Cane, iPhone	4
P6	woman	58	Totally Blind	no	White Cane, iPhone, Guide Dog	4
P7	woman	46	Totally Blind	yes	White Cane, iPhone, Milestone	4
P8	man	27	Totally Blind	yes	Care Taker, Wheelchair, White Cane, iPhone, Glasses	4
P9	man	63	Totally Blind	yes	White Cane, iPhone, Headphones	5
P10	man	52	Totally Blind	yes	White Cane, iPhone	5
P11	woman	37	Low Vision	yes	White Cane, iPhone, Magnifier (16x), Monocular	4
P12	woman	50	Totally Blind	yes	White Cane, iPhone	3
P13	man	42	Low Vision	no	White Cane, iPhone	3

LifeInsight. The accuracy metrics allow us to evaluate how close LifeInsight’s visual answering is compared to a human answer. We used Gonzalez et al. [52]’s definition of accuracy: *“the extent to which the scene description matched a sighted human perception of the image”* [52]. The accuracy metric ranges from 1 to 3, with 1 no match at all, 2 a partial match, and 3 great match. LifeInsight’s accuracy scores are drawn from three researchers’ coding considering the tasks, the visual input, each researcher’s answer given the task and the visual input, and LifeInsight’s response given the visual and auditory input. This allowed us to evaluate the quality of LifeInsight’s output experienced by the participants and to compare it with prior work by Gonzalez et al. [52].

6.2.2 Phase 2: Reflection on Using LifeInsight. We concluded with semi-structured interviews to learn more about the participants’ perceptions and opinions when using LifeInsight, which we transcribed and translated for analysis. The insights help us in shaping the design and interaction interface of future AI-powered assistive wearables. The interviews were guided by (1) usage experience, (2) motivation of using LifeInsight in daily life, and (3) design and interaction space for future AI-powered assistive wearables. Affinity diagramming [43] was used to cluster participants’ feedback, following a similar procedure as outlined in Section 3.4. The lead researcher first introduced the study context and the semi-structured interview script to the other researchers. All researchers were aware of the research but some were less aware of the individual steps and the data collection procedures. The semi-structured interview script was available on the digital whiteboard for the duration of the affinity diagramming. Three researchers then employed affinity diagramming [43] to cluster and organise the interview data. The clustering session lasted for approximately 1 hour and was performed on Miro [4]. Labels were attached towards the end of the clustering, with another round of discussion to resolve any remaining comments and thoughts by the team. Appendix C provides an overview of the core questions.

7 RESULTS: USER STUDY OF LIFEINSIGHT

We first present LifeInsight’s performance in assisting BLV people across the scenarios, followed by insights into usage and querying behaviours. Finally, we report on LifeInsight’s usability and participants’ perceived workloads, their empathy levels when using LifeInsight, and the findings from the interviews.

7.1 Phase 1: Using LifeInsight in Daily Scenarios

7.1.1 Usage Success. Participants could successfully use LifeInsight in the majority of the scenarios (80.8%, 63 out of 78 experiences). While LifeInsight could identify the state of the candle and correctly provide the information shown on the display for all participants, some participants faced challenges when using LifeInsight in the other scenarios, which we will report in the following.

Localising Objects (76.9%). Participants encountered challenges in identifying whether the watch they were searching for was on the table on three occasions. In these situations, LifeInsight did either not accurately describe the watch on the table (e.g., *“a black object”*, P5) or informed the participants wrongly about the absence of the watch (e.g., *“No, I don’t see a watch on the table.”*, P7 and P12). Interestingly, P7 followed up by asking LifeInsight if it sees the entire table, on which LifeInsight replied *“No, the table is not fully visible. Only a part of it can be seen in the foreground on the right”*. P7 made two additional visual queries about the presence of the watch, which is a common method by BLV people to check for conflicting information among multiple model outputs [60]. However, LifeInsight responded negatively both times as the watch was not within LifeInsight’s field of view. Consequently, P7 concluded that there was no watch on the table.

Comparing Objects (76.9%). Three participants faced challenges using LifeInsight to identify which jar contained honey. Specifically, P2 requested LifeInsight to read aloud the text on the jar, but LifeInsight was unable to accurately analyse the taken picture. Similarly, for P3 and P5, LifeInsight provided

incorrect answers based on the visual input. For example, in P5's case, they held a jam jar in their left hand while LifeInsight's camera focused on the honey jar on the table, resulting in incorrect information about what the participant was actually holding in their hand.

Comprehending Text (69.2%). There were four participants who experienced challenges in identifying the appropriate washing temperature of the towel using LifeInsight. For example, P4 was holding the care label of the towel and covered the washing instructions that showed the 60°C information with their fingers. In the case of others, such as P8, the relevant information on the care label was not visible and a *generic* prompt read aloud other visible parts of the label. In the case of P10, LifeInsight provided a response to the participant's visual query but the response was wrong (40°C instead). In the case of P9, the washing label was too blurry for LifeInsight to identify the washing temperature.

Navigating in (un-)familiar Environments (61.5%). While in some scenarios LifeInsight analysed a visual prompt that deviated from a view of the floor in front of the user (e.g., the door, the participants' arms), there were situations where LifeInsight misinterpreted visual information. For example, for P8, instead of detecting the chair or the cable on the floor, LifeInsight informed the participant about a white card or a piece of paper on the floor as the object's were only partially visible by LifeInsight.

7.1.2 Usage Behaviours & Visual Querying Behaviours. While we primarily designed LifeInsight as a wearable that can be clipped onto the collar (default position), we did not restrict the participants' usage behaviours. The majority of the participants kept LifeInsight attached to their collar (in 69 out of 78 experiences, 88.46%). Three participants, P7, P8 and P11, decided to use LifeInsight as a handheld device in some scenarios. For example, P7 took LifeInsight into their left hand to assist in identifying the washing temperature of the towel. P8 held LifeInsight in their left hand when aiming to identify whether there is a watch on the table as this helped them to adjust their visual coverage of the table in front of them.

Overall, participants queried LifeInsight 307 times ($M = 23.62$ prompts per participant; $M = 3.94$ prompts per scenario). Distinguishing two objects ($M = 4.62$, $SD = 3.43$) and comprehending written text ($M = 5.00$, $SD = 4.49$) resulted in the highest number of prompts. Contrary, localising the watch on the table ($M = 2.62$, $SD = 1.19$) and navigating in (un-)familiar environments ($M = 3.38$, $SD = 2.43$) resulted in the lowest number of prompts. Table 6 in Appendix F provides an overview of all scenarios.

When exploring the most frequent querying method, we observed participants' visual queries considered both the object of interest and the visual information queried for, i.e., they requested goal-oriented information [124]. For example, when identifying the state of a candle, most participants queried specifically if the candle is lit (e.g., "Is there a candle burning on the table?"). Across all scenarios, *targeted object-centred*

prompt represented the majority of visual queries, 127 out of 307 requests (41.4%). The second most common querying approach entailed the object of interest without a targeted visual query. For example, when comparing objects, participants asked questions such as "What is written on the jar?" or "What do I hold in my hand?" without directly requesting the specific information they sought (like P2, "Which of the two glasses contains honey?"). This visual querying method covered 30.6% of the prompts, 94 out of 307. The remaining queries fall either into the *generic* category (49 out of 307, 15.9%) or were incomplete due to inaccurate auditory sensing (37 out of 307, 12.1%, e.g., "what" instead of "what do you see?").

Interestingly, we observed scenarios where the participants initially used generic prompts and then refined them to more accurately address the specific visual questions they had. For example, P14 began with a generic question "what is on the table?" and then adjusted it to "is there a watch on the table?" to more specifically query for visual information. Others, such as P4, began with a generic prompt "what do you see?" when aiming to access the washing information on the care label of the towel to then transition through an experience with an *object-centred prompt* to a *targeted object-centred prompt*. After being not successful in retrieving an answer using a generic prompt (e.g., a number of degrees), P4 adjusted their prompting and queried LifeInsight more specifically to read aloud what is written on the care label. Since LifeInsight's response still did not specify the temperature at which the product could be washed, P4 refined their query to more explicitly request the needed information (i.e., "with which temperature can I wash this towel?").

7.1.3 Perceived Workload, Perceived Empathy of LifeInsight and General Perception. Participants' perceived workload when using LifeInsight received a mean score of 22.56 on a scale from 0 to 100 ($SD = 13.33$, $\min = 4.17$, $\max = 46.67$). Table 5 in Appendix E provides an overview of all raw NASA-TLX subscales [59]. For participants' perceived system usability [35], LifeInsight received a score of 74.81 ($Mdn = 80$, $SD = 18.86$), which stands for an above good usability [21]. PETS-ER received a mean score of 48.53 ($SD = 17.67$) and PETS-UT a mean score of 36.83 ($SD = 24.57$). The overall mean score was 43.85 out of 100 ($SD = 19.20$). We compare the PETS scores with the four validation scenarios in the original paper, as suggested by Schmidmaier et al. [104]. LifeInsight showed higher empathy compared to the non-empathic systems by Schmidmaier et al. [104] but lower empathy towards their empathic systems. Furthermore, LifeInsight received a score of $M = 5.15$ ($Mdn = 5.5$, $SD = 1.93$) for satisfaction, $M = 5.60$ ($Mdn = 6$, $SD = 1.67$) for participants' trust in its responses, and $M = 5.26$ ($Mdn = 6$, $SD = 2.05$) for perceived success on 7-Point Likert scales.

7.1.4 LifeInsight's Visual Understanding vs. Human's Visual Understanding. LifeInsight's accuracy across all scenarios was 2.64 on a 3-point scale ($SD = 0.62$), slightly higher compared to Gonzalez et al. [52]'s app with a mean of 1.95 ($SD = 0.82$). This suggests LifeInsight performed considerably well

when providing responses to visual questions. Interestingly, there were scenarios where the performance of LifeInsight was superior than the performance of the visual evaluation of three researchers. For example, when P13 aimed to identify the washing temperature of the towel, none of the researchers was able to derive the temperature information solely from LifeInsight's visual sensing. Contrary, LifeInsight was capable of identifying the object as a towel and then determining the temperature at which it can be washed. This suggests that in situations where humans may face challenges in determining the answer based solely on visual information, LifeInsight can bridge this gap by extracting relevant details from other sources. In this case, the type of household linen, a towel, provided the relevant information. An overview of the images used for the analysis is provided in [Appendix G. Table 6](#) in [Appendix F](#) provides an overview of the accuracies in each scenario.

7.2 Phase 2: Reflection on Using LifeInsight Across Multiple Scenarios

7.2.1 Experience and Perception Using LifeInsight. LifeInsight received overall positive feedback as an AI-powered assistive wearable. Participants voiced *“the simplicity of pressing a button is very good and handy for everyday use”* (P2) and that what sets LifeInsight apart from other ATs (e.g., Seeing AI [7] and Be My Eyes [6]) is *“the speech input and speech output, which [P1] otherwise only knows from Siri”* (P1). Another positive point that stood out was LifeInsight's design, which allowed participants to wear and use it on-body.

“The advantage of this system is that you have your hands free. So if I take something off the shelf in a store, I can easily take the device in one hand and hold my cane in the other. [...] With Be My Eyes, it's practically impossible [...] I have my cane, and I should hold the phone, which is nearly impossible to operate with one hand. [...]” - P13

Furthermore, participants were positive about the versatility of LifeInsight and about its responses to the visual queries, which participants termed as advantageous over existing ATs.

“What I really liked here is the flexibility [...] I was pleasantly surprised. So far, the systems I've tried were more specialised [...] Here, I have a wide range of options for how I can ask and how the system then develops an answer.” - P9

“In the app version, I have to first take the phone, start the app, say what I want, and then I can film or take a photo. And here, you'd have it mounted.” - P7

“On a smartphone, I need to specifically install an app for recognition. And this system could be used for various things, like colour recognition, text recognition, and so on.” - P6

P13 highlighted the capabilities of LifeInsight to perform targeted visual queries and that these affordances accelerate their experiences in daily scenarios compared to existing ATs.

“If I compare the prototype with OrCam or Seeing AI, I find the information was much clearer. I heard much less nonsense. With Seeing AI or the OrCam, when I look at a product, I always have to guess from bits of words what it might be.” - P12

Despite the positive words about LifeInsight, there were some situations where the use of LifeInsight was challenging and resulted in multiple visual queries that negatively impacted participants' experiences. For example, P1 voiced that multiple attempts have reduced satisfaction a bit but that they are used to it from prior experiences using ATs (e.g., Seeing AI [7] or Be My Eyes [6]).

“The problem I just described [camera orientation]: you actually have with many aids. Sometimes you just have to aim precisely to achieve reasonable results.” - P9

Related to P9's comment, P1 voiced that what is still missing is for LifeInsight to better understand what *“[they] want and help optimise; built-in usage aids and optimisation options”* (P1). P12 added that they believe the way the visual question is asked might require a bit more practice.

7.2.2 Motivation Using LifeInsights in Everyday Life. When asked what would motivate the participants to use AI-powered assistive wearables in their daily lives, P8 mentioned that LifeInsight would provide them with *“more satisfaction and the ability to move through life like a sighted person.”* (P8). Others emphasised a reduced dependency on other people.

“It would greatly ease my daily life because I wouldn't have to think so much about where to find a sighted person to ask.” - P12

“It would certainly be more convenient than always using my iPhone. With the iPhone, I have to hold it in one hand and operate it with the other. If you also have a guide dog or a white cane, it would be nice to have a button you can use one-handed.” - P4

Although participants argued that AI-powered assistive wearables could replace human assistance in many situations (e.g., P5: *“the system is available around the clock, which a human usually isn't, even when living in assisted living”*), some participants mentioned scenarios where they would still prefer human assistance.

“If I am alone, I would definitely use the system, or if I don't feel like asking someone. Otherwise, if I have the object in my hand and someone is standing right next to me, I would say 'take a quick look'.” - P6

Participant P12 noted that their motivation for using AI-powered assistive wearables depends on the specific context and the amount of trust they have in such systems.

"I would need a human when I feel it's absolutely important to get a reliable answer that I can trust. If I had a job interview tomorrow and I need to know which clothes are really spotless, I don't think I would trust the system 100%. I would rely on a sighted person." - P12

7.2.3 Perception of LifeInsight's Design. Participants found LifeInsight easy to use and appreciated being in control of the visual querying using a single-click interface.

"With other systems like Seeing AI, the camera just runs and talks whenever it recognises something. Here, I can control it myself when I want information, which I greatly appreciated." - P3

"I like that [you have to press [the button], when I can trigger something myself and not everything always happens automatically. In certain situations, it would be cool, I wouldn't need the hands because they're already occupied, maybe it could be via a code word but not either/or but rather both." - P10

There were some participants who voiced interest in using LifeInsight in combination with a wake word [8].

"I would like to be able to say, 'Hey Prototype, I want this and that.' The problem is, I'm out in the city, and I have the cane in my right hand and the guide dog in my left hand, then I have to let go of something to press the button." - P9

"I would prefer voice control over a button because I often want to use it when my hands are occupied or dirty." - P12

8 DISCUSSION

We introduced and evaluated LifeInsight, an AI-powered assistive wearable for BLV people. The cultural probes provided rich insights into BLV people's daily experiences and played a pivotal role in the concept, design and evaluation of LifeInsight. We now discuss the key findings in relation to the two research questions and conclude by discussing key lessons and practical recommendations for the design, development, and evaluation of future AI-powered assistive wearables for BLV people.

8.1 RQ₁: To what extent can an AI-powered assistive wearable address BLV people's needs and preferences in multiple everyday life scenarios?

The use of LifeInsight resulted in a low perceived workload and an "above good" [21] usability score. LifeInsight assisted BLV people in multiple situations that, as reported by the participants, would otherwise be challenging for them, or even close to impossible. Furthermore, participants valued the goal-oriented approach to their visual queries which, in comparison to existing ATs such as Seeing AI [7], reduced the amount of general descriptions of the real-world surroundings that

do not provide specific answers to their objectives. Although LifeInsight still faced challenges for accurately querying visual information in some scenarios (Section 7.1.1 and Section 7.2.1), it worked remarkably well across the majority of the scenarios. As suggested by LifeInsight's evaluation, argued by Kane et al. [69], and implemented by Bigham et al. [26] with VizWiz, ATs "equipped with AI" - be it through multimodal LLMs like in LifeInsight or simulated through crowd-workers like in VizWiz [26] - can indeed reduce the amount of carrying around multiple ATs. As deemed promising future research by Gonzalez et al. [52], LifeInsight's evaluation provides empirical evidence that integrating multiple visual querying affordances into one holistic experience has a positive impact on the usability of ATs. While LifeInsight accommodates multiple daily scenarios, its design and interaction was still perceived as highly usable. In fact, LifeInsight was perceived as superior to existing ATs, such as Be My Eyes [6], which is used in more than 150 countries by more than 650,000 BLV people [6]. We observed scenarios where LifeInsight surpassed human's performance when evaluating visual content, such as the "towel example" in Section 7.1.4. This finding is particularly promising as it suggests that LifeInsight can deliver accurate responses to BLV people's visual queries without direct visual access to the information. By accurately identifying specific objects (e.g., a towel), LifeInsight provides empirical evidence of its potential to replicate certain aspects of human cognitive abilities. Some participants assumed the ideal washing temperature to be 60°C based on identifying the object as a towel by touch. In these situations, LifeInsight can still be used to validate BLV people's initial assumptions, which becomes especially valuable in more hazardous scenarios, such as when checking the status of an indoor fireplace (Figure 3). Despite LifeInsight's assistance and effectiveness, it is important to acknowledge that in scenarios like the "towel example", BLV people, caretakers, and others could provide similar information through a tactile inspection of the towel and their knowledge of textile care and appropriate washing temperatures.

Answer to RQ₁

AI-powered assistive wearables can assist BLV people in multiple scenarios. Compared to existing assistance, such as Seeing AI [7] or human support, LifeInsight offers notable advantages. First, LifeInsight's wearable form factor reduces friction as BLV people do not need to grab an external device or navigate to the relevant assistive app on their smartphone. Second, while human assistance is not always available, LifeInsight offers a reliable alternative, enhancing BLV people's independence by minimising (or even eliminating) the need to rely on another person's constant availability.

8.2 RQ₂: How do BLV people interact with an AI-powered assistive wearable across multiple everyday life scenarios?

When exploring participants' querying behaviours, the majority of the queries can be classified as targeted visual queries, where the user queries for specific visual information (i.e., goal-oriented [124]). This suggests LifeInsight allows for assistance that is tailored to BLV people's visual queries instead of providing generic visual scene descriptions. Interestingly, P12, for example, voiced they found the information from LifeInsight much clearer than from Seeing AI [7] or the OrCam [115] as they heard "*much less nonsense*" and do not have to "*guess from bits of words what it might be*" (P12's quote in Section 7.2.1). The querying behaviours align with the importance of future applications that "*tailor their output through settings and instructions from the user*" and allow for continuous conversations [52]. By combining multimodal LLMs with speech input, as implemented in LifeInsight, participants adjusted their input for more detailed output through their natural use of LifeInsight. For example, P4 began with a generic query when aiming to access the washing information on the care label to then realise their query does not yield an output that is specific enough. Accordingly, they adjusted their query to receive a more specific output. This suggests that compared to non-conversational ATs [7, 74, 93], LifeInsight allows BLV people to reflect on, and verify, the received responses through follow-up queries, an important ability for obtaining additional information when needed by BLV people [60, 116]. Furthermore, P4's experience highlights how LifeInsight allows starting with generics and moving into specifics to "*tactilely explore photo 'layers' for specific information*" [52]. This concept has already demonstrated promising results for BLV people in multi-layered touch explorations on images [74]. LifeInsight provides additional empirical evidence of its effectiveness, showing how AI-powered assistive wearables can provide meaningful assistance across various scenarios in BLV people's everyday lives.

While the concept and design of LifeInsight offer significant flexibility, questions about user adaptations remain open, particularly highlighted by P4's example, where they refined their input to obtain a more accurate response (Section 7.1.2). We believe that future research on AI-powered assistive wearables should incorporate BLV people's behaviors and preferences into the AI-powered processing pipeline as best as possible. For example, aligning with the work by Mott et al. [92], which adapted touch interfaces for individuals with motor impairments, wearable ATs for BLV people could be further enhanced using teachable AI [91, 121] or incorporating accessibility data into AI model training [3]. These approaches allow BLV people to personalise their experiences based on their needs and interaction behaviours, e.g., when performing visual queries, potentially minimising the amount of additional user adaptations required. Partnerships between Microsoft and Be My Eyes [3], which focus on incorporating accessibility data into AI model training, already undertake a significant step forward to better address BLV people's needs

by "*making AI models more inclusive for the over 340 million people around the world who are blind or have low vision*" [3]. Such efforts, if done in a privacy-preserving manner [2], can reduce AI biases and AI errors [16], as well as better tailor the output of AI-powered assistive wearables to BLV people's needs and prompting behaviours.

Answer to RQ₂

Participants employed three methods when querying visual information: *generic prompts*, *object-centred prompts* and *targeted object-centred prompt*, with *targeted object-centred prompts* representing the most frequent one applied. Despite LifeInsight's novel design, featuring a single-click interface combined with an AI-powered assistive wearable, participants encountered minimal interaction friction and reported finding LifeInsight highly natural to use. While most participants wore LifeInsight attached to their collar, some preferred to use it as a handheld device in certain scenarios.

8.3 Key Lessons, Future Research Directions, and Practical Recommendations for Designing AI-Powered Assistive Wearables for BLV People

Over the past 1.5 years, we distilled three key lessons based on the cultural probes and when designing, implementing and evaluating LifeInsight for the first time. These insights are intended to aid researchers and practitioners in fostering a shared understanding of the collaborative development of future AI-powered assistive wearables that cater to BLV people's needs and preferences.

8.3.1 Key Lesson 1: Identifying Challenges and Contributing Assistance to BLV People's Daily Experiences: Less Specialised, More Multifunctional. Through the cultural probes, we identified existing challenges faced by BLV people in their everyday lives and how BLV people envision future ATs that accommodate their needs and preferences. While the results from the probes are not the first that contribute a taxonomy of BLV people's experiences in their daily lives (e.g., [32, 80, 102]), from our work with BLV people we learned:

- about a potential promising direction for future AI-powered assistive wearables that encompasses the current needs and preferences of BLV people across multiple daily scenarios, and
- how to design a study around common daily scenarios of BLV people to provide empirical insights into, and evidence of, the use of AI-powered assistive wearables.

Reflecting on the probes, we learned that BLV people have a desire for ubiquitous access to ATs that are integrated into their daily life unobtrusively, akin to the presence of assistive mobile devices (e.g., an iPhone + VoiceOver⁴). As

⁴<https://www.apple.com/accessibility/vision>, last accessed 6 Aug. 2024

noted by the participants, such an integration of assistance can offer support even in unforeseen challenging situations, sparing the burden of anticipating potential obstacles later in the day and carrying around multiple ATs. As articulated by Zolyomi [132], “*creating accessible HCI requires deep knowledge of disabled people and the current state of accessible technology*”. By first identifying the daily experiences of BLV people and then using these insights to inform the evaluation of LifeInsight, we contribute to the vision of assistance that is universal, rich and supportive in various daily scenarios. We argue that future research is required to embed LifeInsight into real-world scenarios over several months to help the HCI and accessibility communities further enhance AI-powered assistive wearables and better understand BLV people’s potential behavioural changes over time. This includes examining how they might adjust their visual querying behaviours using AI-powered assistive wearables, as well as how camera-based wearables provide answers to visual questions w.r.t BLV people’s preferences and functional roles/information types [61].

8.3.2 Key Lesson 2: Navigating the Promises and Challenges of AI-Powered Assistive Wearables. With LifeInsight, we were able to provide empirical insights into an AI-powered assistive wearable for BLV people. Participants commended LifeInsight for its versatility, noting that it enabled them to use one AT with a single-click interface and speech input to seamlessly access assistance across various scenarios. They also highlighted its hands-free, lightweight and wearable form factor. However, there were concerns about LifeInsight’s field of view and potential privacy concerns of camera-based ATs when used daily (confirming previous survey-based research by Akter et al. [15]). First, LifeInsight’s evaluation showed that participants prefer to maintain control over the visual sensing, including what and when the system processes information. In this context, research is encouraged to explore alternative activation methods (e.g., wake words [8]) and to assist BLV people in positioning LifeInsight’s camera for visual sensing (Section 7.2.1 and Section 7.2.3). For example, from neighbouring fields we can learn how to embed alternative camera configurations [40] into future AI-powered assistive wearables, such as to achieve stereo vision [127] and provide an understanding of the visual context around the user [126]. Second, the literature emphasised the importance of developing privacy-preserving ecosystems for camera-based wearables [15, 22] and AI systems [46]. While LifeInsight incorporates measures to protect privacy, such as the integration of an additional privacy layer (Figure 4), research is needed that focuses on enhancing LifeInsight’s ability to preserve BLV people’s and bystander’s privacy [14] while identifying the optimal balance with its visual sensing capabilities (and assisting BLV people in taking high-quality pictures [64]). As indicated by P12 in Section 7.2.1, it remains important to learn more about how BLV people query AI-powered assistive wearables when they ask visual questions, and to use these insights to enhance the training of AI models, which we will discuss in more detail in Section 8.3.4.

8.3.3 Key Lesson 3: The Necessity of Human-Centred Design in AI-Powered Assistive Wearables for BLV People.

AI that guides humans through everyday life has for most people “*mostly been confined to science fiction*” [24] and, according to Bigham and colleagues [24, 25], “*people with disabilities have been using systems with many of these features for decades, placing people with disabilities at the forefront of artificial intelligence*”. While it is essential to consider i) “*special care to ensure that systems are trained and evaluated on the diversity of abilities and experiences that people with disabilities have*” [66], and ii) providing custom labels from BLV people [66, 121], we suggest that there may be other factors that could further enhance the development of effective AI-powered assistive wearables. For example, in widely used applications such as Seeing AI [7], it is essential to acknowledge that their usefulness is contingent on BLV people having access to a smartphone that meets their requirements. Learning to use a smartphone is a long process where resources are potentially inaccessible and isolating [45]. Our evaluation suggests that BLV people face little to no challenges when using an AI-powered assistive wearable that encompasses a single-click interface and speech input. As noted in Gamage et al. [49]’s literature review, BLV people prefer voice input and speech output when interacting with ATs, both of which are incorporated into LifeInsight. Despite P8’s experience when using a head-mounted AT in public (Section 4.2), our findings as part of LifeInsight’s empirical evaluation suggest that future ATs may not necessarily need to “*strive to be designed for social acceptability*” [108–111]. Participants highlighted the importance of functionality, especially as technology becomes increasingly ubiquitous in public and social settings. The little evidence for *social acceptance* in our study might be attributed to several factors: i) the time when Shinohara and Wobbrock [110]’s research was first published, which predates the current widespread presence of technology in public and social settings, ii) differences in research methods, such as the surveys used in [110] compared to the diary studies in [111] and our in-the-wild probes and laboratory study of a novel wearable AT, or (iii) the design of LifeInsight, which aimed to be discreet and avoid “*marking [blind people] as being blind within a predominantly sighted social world*” [109].

However, in line with Shinohara et al.’s works, we believe it is now time for future research to find clarifying answers to these questions in the context of AI-powered assistive wearables (e.g., by applying the Design for Social Accessibility (DSA) perspective and including the DSA Method Cards [112] in future research). Additionally, the findings encourage future research to explore how AI-powered assistive wearables can further enhance the information details for BLV people by, for example, i) allowing them to personalise the experience for their own needs using teachable AI [91, 121], ii) going “*beyond domain-agnostic factual accuracy to consider how to deliver relevant information efficiently*” [61], and iii) adding certainties or uncertainties [72, 125] to LifeInsight’s output to allow BLV people to independently determine the quality of AI-powered responses and inform them about potential system-introduced AI hallucinations [124].

AI-powered assistive wearable alone, which we consider as today's "cutting edge technology" [110], are insufficient without comprehensive empirical evaluations, including how BLV people prefer to wear them, use them for visual querying, and foster social interactions [90, 111]. We argue future research on AI-powered assistive wearable must also include policy discussions around supporting BLV people in building "accurate mental models, verifying information accuracy, and combating biases and harms" [12]. Integrating these factors, including "cutting-edge technology" [110] and social acceptability constructs [111, 112], are essential for the design and implementation of future AI-powered assistive wearables with BLV people.

8.3.4 Practical Recommendations for Designing AI-Powered Assistive Wearables for BLV People. Based on our empirical research and the key lessons, we synthesise our work into three practical recommendations to act as a resource and reference for researchers and practitioners when contributing to future works on AI-powered assistive wearables for BLV people.

- (I) **Incorporate Wearability, Flexibility and Unobtrusiveness.** We recommend to design future ATs with wearability, flexibility, and unobtrusiveness in mind. The use of a body-mounted interface (i.e., camera and single-click interface) combined with natural speech-based input [48, 50], as implemented in LifeInsight, has shown significant promise across various daily experiences of BLV people. The empirical insights presented in this paper support the value of "hands-free accessibility" for engaging across multiple daily scenarios (e.g., as pushed forward through the partnership between Meta and Be My Eyes [1]). However, our findings, such as P8's comment on OrCam [115] in Section 3.4.2, or the participants' usage behaviours of LifeInsight (Section 7.1.2), challenge the anticipated form factor of future AI-powered assistive wearables. Ideally, BLV people should benefit from user-centred cutting-edge technology, such as AI, while maintaining the ability to customise the look and feel of the final product, such as the form factor of wearable ATs. Depending on BLV people's preferences and needs, this could take on the form of an unobtrusive on-body device, similar to LifeInsight, or of an integration into glasses, such as in the example of the Ray-Ban Meta Smart Glasses [1].
- (II) **Find the Balance Between User Adaptations and System Design.** Whenever feasible, AI-powered assistive wearables for BLV people should aim to better understand BLV people's needs and, as highlighted by P1 in Section 7.2.1, incorporate built-in usage aids to support them in optimising their interactions. The interaction with future AI-powered assistive wearables should remain adaptable by BLV people, similar to LifeInsight, which can be worn on the body or used as a handheld device. As evidenced by P4's visual querying example in Section 7.1.2, we recommend to integrate the challenges and experiences of BLV people into the training process

of AI models, as exemplified by the recent collaboration between Microsoft and Be My Eyes [2] or by the use of teachable AI [91]. While this approach should carefully balance the additional effort required for personalisation with the benefits it provides to BLV people [91], these adaptations can be highly beneficial for frequent and long-lasting use of AI-powered assistive wearables. For example, when AI models are trained for specific applications (e.g., assisting BLV people in particular contexts), additional user adaptations, such as those described by P4, may become redundant. The message here is that these user adaptations could potentially be minimised - or even eliminated - as multimodal AI models may become more effective at processing BLV people's input by utilising prior visual queries and long-term usage behaviours.

- (III) **Align the Design of Wearable ATs with BLV People's Needs and Existing (Analog) Support Systems.** While numerous technological challenges remain, such as delays in AI-based systems due to the handling of complex visual input and speech input, we believe that now is the right time to collaborate towards a future where AI-powered assistive wearables "make our interactions with computers, the world around us, and other people more accessible to everyone" [24]. We recommend that future works prioritise continuous learning about BLV people's everyday scenarios, preferences, and social contexts - including how future AI-powered assistive wearables complement existing more traditional, non-digital systems (e.g., P13's comment in Section 7.2.1 and P4's comment in Section 7.2.2). As illustrated by LifeInsight's design decision to prioritise freeing up the hands of BLV people, we recommend future explorations on how AI-powered assistive wearables can complement established support systems. This is particularly important given that existing popular aids like the white cane serve dual functions: they not only assist BLV people to navigate safely but also affect the BLV people's surroundings as they make other people aware of the disability and remind them "to leave space and show respect" [20]. While existing powerful ATs (e.g., Vocatex Plus [10]) offer significant assistance in BLV people's lives, there is still a pressing need for future systems that can be seamlessly and consistently integrated into everyday life, particularly in situations where human assistance is unavailable and where BLV people experience dynamic and mobile situations (e.g., while walking). Inspired by Bigham's quote that "[...] just because something is impressive does not mean it will be adopted" [24], the recommendation is not to reinvent the wheel but to explore how technology can be more effectively and seamlessly embedded into the lived experiences of BLV people, which we aimed to achieve by transitioning from early promising smartphone-based apps, like VizWiz [26], to AI-powered assistive wearables, like LifeInsight. Drawing on research with multiple BLV people over a significant

period, it remains important to align the design and interactional affordances of future AI-powered assistive wearables with existing aids and coexist with them, fostering a more comprehensive and impactful ecosystem of assistance for BLV people.

9 CONCLUSION

In this paper, we presented and evaluated LifeInsight, an AI-powered assistive wearable for BLV people. Using cultural probes and interviews, we first contributed an understanding of the daily experiences of BLV people. By collecting and analysing empirical data acquired from 10 BLV people in real-world settings, we not only raised awareness of current challenges but also informed the design, development and evaluation of LifeInsight. We evaluated LifeInsight across six scenarios with 13 BLV people. Our work highlighted the potential of AI-powered assistive wearables for BLV people, identified BLV people's visual querying behaviours, and outlined promising future research directions for the development and evaluation of AI-powered assistive wearables. With LifeInsight, including the cultural probes and LifeInsight's design, implementation and empirical evaluation, we make meaningful contributions to a human-centred future that encompasses suitable assistance in various daily situations experienced by BLV people.

ACKNOWLEDGMENTS

We want to thank all our collaboration partners for helping in the participant recruitment, particularly the Swiss National Association of and for the Blind (SNABLIND), Schweizerischer Blinden- und Sehbehindertenverband (SBV, Sektion Ostschweiz), and Lebenshilfe Vorarlberg. We also want to thank SIGCHI and the internal funding from the University of St. Gallen for their instrumental role in realising the CHI Together event [105]. We thank all reviewers for their valuable comments on our work, and all participants for taking part in our research. Finally, thank you to Lydia M. and Carolin S. for helping with the teaser figure. This research was supported by the International Postdoctoral Fellowships (IPF, University of St. Gallen). In accordance with ACM's policy⁵ on the use of generative AI during the paper preparation process we want to disclose the following. In this paper, we used Overleaf's built-in spell checker, Grammarly, and the current version of ChatGPT (January 2025). These tools helped us fix spelling mistakes and get suggestions to improve the writing of the paper.

REFERENCES

- [1] 2024. *Be My Eyes and Meta Announce Accessibility Partnership*. Retrieved November 18, 2024 from <https://www.bemyeyes.com/blog/be-my-eyes-meta-accessibility-partnership>
- [2] 2024. *Be My Eyes Launches Initiative to Make AI Models More Inclusive*. Retrieved November 18, 2024 from <https://www.bemyeyes.com/blog/be-my-eyes-launches-initiative-to-make-ai-models-more-inclusive>
- [3] 2024. *Microsoft to Help Train Inclusive AI Models Be My Eyes Announces New Collaboration with Microsoft to Help Train Inclusive AI Models*. Retrieved November 18, 2024 from <https://www.bemyeyes.com/blog/be-my-eyes-announces-new-collaboration-with-microsoft>
- [4] 2024. *Miro: The Visual Collaboration Platform for Every Team*. Retrieved November 18, 2024 from <https://miro.com>
- [5] 2024. *OpenAI's GPT-4 Vision API*. Retrieved April 10, 2024 from <https://platform.openai.com/docs/guides/vision>
- [6] 2024. *See the world together*. Retrieved Feb 08, 2024 from <https://www.bemyeyes.com/>
- [7] 2024. *Seeing AI - Talking Camera for the Blind*. Retrieved Feb 08, 2024 from <https://www.seeingai.com/>
- [8] 2024. *Siri, Apple*. Retrieved July 09, 2024 from <https://www.apple.com/siri/>
- [9] 2024. *Speech recognition engine*. Retrieved April 10, 2024 from <https://pypi.org/project/SpeechRecognition/>
- [10] 2024. *Vocatex Plus 4*. Retrieved November 18, 2024 from <https://accesstech.ch/produkt/vocatex-plus-hd/>
- [11] Ali Abdolrahmani, William Easley, Michele Williams, Stacy Branhman, and Amy Hurst. 2017. Embracing Errors: Examining How Context of Use Impacts Blind Individuals' Acceptance of Navigation Aid Errors. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) (CHI '17). Association for Computing Machinery, New York, NY, USA, 4158–4169. <https://doi.org/10.1145/3025453.3025528>
- [12] Rudaiba Adnin and Maitraye Das. 2024. "I look at it as the king of knowledge": How Blind People Use and Understand Generative AI Tools. *ASSETS '24* 16, 54 (2024), 92.
- [13] Dragan Ahmetovic, Federico Avanzini, Adriano Barate, Cristian Bernareggi, Marco Ciardullo, Gabriele Galimberti, Luca A. Ludovico, Sergio Mascetti, and Giorgio Presti. 2023. Sonification of navigation instructions for people with visual impairment. *International Journal of Human-Computer Studies* 177 (2023), 103057. <https://doi.org/10.1016/j.ijhcs.2023.103057>
- [14] Taslima Akter, Tousif Ahmed, Apu Kapadia, and Swami Manohar Swaminathan. 2020. Privacy Considerations of the Visually Impaired with Camera Based Assistive Technologies: Misrepresentation, Impropriety, and Fairness. In *Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility* (Virtual Event, Greece) (ASSETS '20). Association for Computing Machinery, New York, NY, USA, Article 32, 14 pages. <https://doi.org/10.1145/3373625.3417003>
- [15] Taslima Akter, Bryan Dosono, Tousif Ahmed, Apu Kapadia, and Bryan Semaan. 2020. "I am uncomfortable sharing what I can't see": Privacy Concerns of the Visually Impaired with Camera Based Assistive Applications. In *29th USENIX Security Symposium (USENIX Security 20)*. 1929–1948.
- [16] Rahaf Alharbi, Pa Lor, Jaylin Herskovitz, Sarita Schoenebeck, and Robin N. Brewer. 2024. Misfitting With AI: How Blind People Verify and Contest AI Errors. In *Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility* (St. John's, NL, Canada) (ASSETS '24). Association for Computing Machinery, New York, NY, USA, Article 61, 17 pages. <https://doi.org/10.1145/3663548.3675659>
- [17] Saki Asakawa, João Guerreiro, Daisuke Sato, Hironobu Takagi, Dragan Ahmetovic, Desi Gonzalez, Kris M. Kitani, and Chieko Asakawa. 2019. An Independent and Interactive Museum Experience for Blind People. In *Proceedings of the 16th International Web for All Conference* (San Francisco, CA, USA) (W4A '19). Association for Computing Machinery, New York, NY, USA, Article 30, 9 pages. <https://doi.org/10.1145/3315002.3317557>
- [18] Shiri Azenkot and Nicole B. Lee. 2013. Exploring the use of speech input by blind people on mobile devices. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility* (Bellevue, Washington) (ASSETS '13). Association for Computing Machinery, New York, NY, USA, Article 11, 8 pages. <https://doi.org/10.1145/2513383.2513440>
- [19] Shiri Azenkot, Kyle Rector, Richard Ladner, and Jacob Wobbrock. 2012. PassChords: secure multi-touch authentication for blind people. In *Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility* (Boulder, Colorado, USA) (ASSETS '12). Association for Computing Machinery, New York, NY, USA, 159–166. <https://doi.org/10.1145/2384916.2384945>
- [20] Maria Bäckman. 2023. Living with an iconic aid: The case of the white cane. *Culture Unbound* 15, 2 (2023).

⁵<https://www.acm.org/publications/policies/frequently-asked-questions>, last accessed 5 Sep. 2024

- [21] Aaron Bangor, Philip Kortum, and James Miller. 2009. Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of usability studies* 4, 3 (2009), 114–123.
- [22] Divyanshu Bhardwaj, Alexander Ponticello, Shreya Tomar, Adrian Dabrowski, and Katharina Kromholz. 2024. In Focus, Out of Privacy: The Wearer's Perspective on the Privacy Dilemma of Camera Glasses. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '24). Association for Computing Machinery, New York, NY, USA, Article 577, 18 pages. <https://doi.org/10.1145/3613904.3642242>
- [23] Palavi V. Bhole, Ziming Li, Shivang Bokolia, Tae Oh, Garreth W. Tigwell, and Roshan L Peiris. 2024. Haptic2FA: Haptics-Based Accessible Two-Factor Authentication for Blind and Low Vision People. *Proc. ACM Hum.-Comput. Interact.* 8, MHCI, Article 264 (Sept. 2024), 20 pages. <https://doi.org/10.1145/3676509>
- [24] Jeffrey P. Bigham. 2024. Artificial Intelligence and Machine Learning + Accessibility. In *Teaching Accessible Computing*, Amy Oleson, Andrew J. Ko, and Richard Ladner (Eds.). <https://bookish.press/tac> Retrieved 13 Nov, 2024.
- [25] Jeffrey P Bigham and Patrick Carrington. 2018. Learning from the front: People with disabilities as early adopters of AI. *Proceedings of the 2018 HCIC Human-Computer Interaction Consortium* (2018).
- [26] Jeffrey P. Bigham, Chandrika Jayant, Hanjie Ji, Greg Little, Andrew Miller, Robert C. Miller, Robin Miller, Aubrey Tatarowicz, Brandyn White, Samuel White, and Tom Yeh. 2010. VizWiz: nearly real-time answers to visual questions. In *Proceedings of the 23rd Annual ACM Symposium on User Interface Software and Technology* (New York, New York, USA) (UIST '10). Association for Computing Machinery, New York, NY, USA, 333–342. <https://doi.org/10.1145/1866029.1866080>
- [27] Jeffrey P. Bigham, Chandrika Jayant, Andrew Miller, Brandyn White, and Tom Yeh. 2010. VizWiz::LocateIt - enabling blind people to locate objects in their environment. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*. 65–72. <https://doi.org/10.1109/CVPRW.2010.5543821>
- [28] M. Billingshurst and T. Starner. 1999. Wearable devices: new ways to manage information. *Computer* 32, 1 (1999), 57–64. <https://doi.org/10.1109/2.738305>
- [29] Roger Boldu, Denys J.C. Matthies, Haimo Zhang, and Suranga Nanayakkara. 2020. AiSee: An Assistive Wearable Device to Support Visually Impaired Grocery Shoppers. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 4, Article 119 (Dec. 2020), 25 pages. <https://doi.org/10.1145/3432196>
- [30] Andy Boucher, Dean Brown, Liliana Ovalle, Andy Sheen, Mike Vanis, William Odom, Doenja Oogjes, and William Gaver. 2018. TaskCam: Designing and Testing an Open Tool for Cultural Probes Studies. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (CHI '18). Association for Computing Machinery, New York, NY, USA, 1–12. <https://doi.org/10.1145/3173574.3173645>
- [31] James V. Bradley. 1958. Complete Counterbalancing of Immediate Sequential Effects in a Latin Square Design. *J. Amer. Statist. Assoc.* 53, 282 (1958), 525–528. <https://doi.org/10.1080/01621459.1958.10501456>
- [32] Erin Brady, Meredith Ringel Morris, Yu Zhong, Samuel White, and Jeffrey P. Bigham. 2013. Visual challenges in the everyday lives of blind people. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) (CHI '13). Association for Computing Machinery, New York, NY, USA, 2117–2126. <https://doi.org/10.1145/2470654.2481291>
- [33] Daniella Briotto Faustino and Audrey Girouard. 2018. Understanding Authentication Method Use on Mobile Devices by People with Vision Impairment. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility* (Galway, Ireland) (ASSETS '18). Association for Computing Machinery, New York, NY, USA, 217–228. <https://doi.org/10.1145/3234695.3236342>
- [34] Anke Brock and Christophe Jouffrais. 2015. Interactive audio-tactile maps for visually impaired people. *SIGACCESS Access. Comput.* 113 (nov 2015), 3–12. <https://doi.org/10.1145/2850440.2850441>
- [35] John Brooke et al. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* (1996).
- [36] Erin Buehler, Stacy Branham, Abdullah Ali, Jeremy J. Chang, Megan Kelly Hofmann, Amy Hurst, and Shaun K. Kane. 2015. Sharing is Caring: Assistive Technology Designs on Thingiverse. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) (CHI '15). Association for Computing Machinery, New York, NY, USA, 525–534. <https://doi.org/10.1145/2702123.2702525>
- [37] Shaojun Cai, Ashwin Ram, Zhengtai Gou, Mohd Alqama Wasim Shaikh, Yu-An Chen, Yingjia Wan, Kotaro Hara, Shengdong Zhao, and David Hsu. 2024. Navigating Real-World Challenges: A Quadraped Robot Guiding System for Visually Impaired People in Diverse Environments. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '24). Association for Computing Machinery, New York, NY, USA, Article 44, 18 pages. <https://doi.org/10.1145/3613904.3642227>
- [38] Ruei-Che Chang, Yuxuan Liu, and Anhong Guo. 2024. World-Scribe: Towards Context-Aware Live Visual Descriptions. In *Proceedings of the Annual Symposium on User Interface Software and Technology* (UIST '24).
- [39] Kathy Charmaz. 2014. *Constructing grounded theory*. sage.
- [40] Ashley Colley, Wouter Van Vlaenderen, Johannes Schöning, and Jonna Häkklä. 2016. Changing the camera-to-screen angle to improve AR browser usage. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*. 442–452.
- [41] GeeksforGeeks | A computer science portal for geeks. 2023. *How to Blur Faces in Images using OpenCV in Python?* Retrieved April 11, 2024 from <https://www.geeksforgeeks.org/how-to-blur-faces-in-images-using-opencv-in-python/>
- [42] Adele Crudden, Michele Capella McDonnall, and Anne Hierholzer. 2015. Transportation: An Electronic Survey of Persons who Are Blind or Have Low Vision. *Journal of Visual Impairment & Blindness* 109, 6 (2015), 445–456. <https://doi.org/10.1177/0145482X1510900603> arXiv:<https://doi.org/10.1177/0145482X1510900603>
- [43] Aby Rikke Friis Dam and Teo Yu Siang. 2023. *Affinity Diagrams: How to Cluster Your Ideas and Reveal Insights*. Retrieved January 30, 2024 from <https://www.interaction-design.org/literature/article/affinity-diagrams-learn-how-to-cluster-and-bundle-ideas-and-facts>
- [44] Pierre Nicolas Durette. 2023. *gTTS (Google Text-to-Speech)*. Retrieved October 24, 2023 from <https://pypi.org/project/gTTS/>
- [45] Isabela Figueira, Yoonha Cha, Jun Zhu, and Stacy M Branham. 2023. Smartphone Stories: Experiences of Blind and Low Vision Older Adults in Acquiring a Smartphone. In *Companion Publication of the 2023 ACM Designing Interactive Systems Conference* (Pittsburgh, PA, USA) (DIS '23 Companion). Association for Computing Machinery, New York, NY, USA, 156–159. <https://doi.org/10.1145/3563703.3596639>
- [46] Leah Findlater, Steven Goodman, Yuhang Zhao, Shiri Azenkot, and Margot Hanley. 2020. Fairness issues in AI systems that augment sensory abilities. *SIGACCESS Access. Comput.* 125, Article 8 (mar 2020), 1 pages. <https://doi.org/10.1145/3386296.3386304>
- [47] Jon E Froehlich, Anke M Brock, Anat Caspi, João Guerreiro, Kotaro Hara, Reuben Kirkham, Johannes Schöning, and Benjamin Tannert. 2019. Grand challenges in accessible maps. *interactions* 26, 2 (2019), 78–81.
- [48] Limin Paul Fu, James Landay, Michael Nebeling, Yingqing Xu, and Chen Zhao. 2018. Redefining Natural User Interface. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) (CHI EA '18). Association for Computing Machinery, New York, NY, USA, 1–3. <https://doi.org/10.1145/3170427.3190649>
- [49] Bhanuka Gamage, Thanh-Toan Do, Nicholas Seow Chiang Price, Arthur Lowery, and Kim Marriott. 2023. What do Blind and Low-Vision People Really Want from Assistive Smart Devices? Comparison of the Literature with a Focus Study. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility* (ASSETS '23). Association for Computing Machinery, New York, NY, USA, Article 30, 21 pages. <https://doi.org/10.1145/3597638.3608955>
- [50] Francisco J García-Peñalvo and Lourdes Moreno. 2019. Special issue on exploring new Natural User Experiences. *Universal Access in the Information Society* 18, 1 (2019), 1–2.
- [51] Bill Gaver, Tony Dunne, and Elena Pacenti. 1999. Design: cultural probes. *interactions* 6, 1 (1999), 21–29.

- [52] Ricardo Gonzalez, Jazmin Collins, Shiri Azenkot, and Cynthia Bennett. 2024. Investigating Use Cases of AI-Powered Scene Description Applications for Blind and Low Vision People. *arXiv preprint arXiv:2403.15604* (2024).
- [53] Timo Götzelmann. 2018. Visually Augmented Audio-Tactile Graphics for Visually Impaired People. *ACM Trans. Access. Comput.* 11, 2, Article 8 (jun 2018), 31 pages. <https://doi.org/10.1145/3186894>
- [54] Jens Grubert, Tobias Langlotz, Stefanie Zollmann, and Holger Regenbrecht. 2017. Towards Pervasive Augmented Reality: Context-Awareness in Augmented Reality. *IEEE Transactions on Visualization and Computer Graphics* 23, 6 (2017), 1706–1724. <https://doi.org/10.1109/TVCG.2016.2543720>
- [55] Yves Guiard. 1987. Asymmetric division of labor in human skilled bimanual action: The kinematic chain as a model. *Journal of motor behavior* 19, 4 (1987), 486–517.
- [56] Yves Guiard. 1988. The kinematic chain as a model for human asymmetrical bimanual cooperation. In *Advances in Psychology*. Vol. 55. Elsevier.
- [57] Anhong Guo, Xiang 'Anthony' Chen, Haoran Qi, Samuel White, Suman Ghosh, Chieko Asakawa, and Jeffrey P. Bigham. 2016. VizLens: A Robust and Interactive Screen Reader for Interfaces in the Real World. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (Tokyo, Japan) (*UIST '16*). Association for Computing Machinery, New York, NY, USA, 651–664. <https://doi.org/10.1145/2984511.2984518>
- [58] Danna Gurari, Qing Li, Abigale J. Stangl, Anhong Guo, Chi Lin, Kristen Grauman, Jiebo Luo, and Jeffrey P. Bigham. 2018. VizWiz Grand Challenge: Answering Visual Questions From Blind People. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- [59] Sandra G Hart. 2006. NASA-task load index (NASA-TLX); 20 years later. In *Proc. of the human factors and ergonomics society annual meeting*. Sage publications Sage CA: Los Angeles, CA.
- [60] Mina Huh, Yi-Hao Peng, and Amy Pavel. 2023. GenAssist: Making Image Generation Accessible. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (San Francisco, CA, USA) (*UIST '23*). Association for Computing Machinery, New York, NY, USA, Article 38, 17 pages. <https://doi.org/10.1145/3586183.3606735>
- [61] Mina Huh, Fangyuan Xu, Yi-Hao Peng, Chongyan Chen, Hansika Murugu, Danna Gurari, Eunsol Choi, and Amy Pavel. 2024. Long-Form Answers to Visual Questions from Blind and Low Vision People. *arXiv:2408.06303* [cs.CL] <https://arxiv.org/abs/2408.06303>
- [62] Liz Jackson, Alex Haagaard, and Rua Williams. 2022. Disability Dongle. <https://blog.castac.org/2022/04/disability-dongle/> last accessed, Aug. 2024.
- [63] Gaurav Jain, Basel Hindi, Zihao Zhang, Koushik Srinivasula, Mingyu Xie, Mahshid Ghasemi, Daniel Weiner, Sophie Ana Paris, Xin Yi Therese Xu, Michael Malcolm, Mehmet Turkcan, Javad Ghaderi, Zoran Kostic, Gil Zussman, and Brian A. Smith. 2024. StreetNav: Leveraging Street Cameras to Support Precise Outdoor Navigation for Blind Pedestrians. In *Proceedings of the Annual Symposium on User Interface Software and Technology (UIST '24)*.
- [64] Chandrika Jayant, Hanjie Ji, Samuel White, and Jeffrey P. Bigham. 2011. Supporting blind photography. In *The Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility* (Dundee, Scotland, UK) (*ASSETS '11*). Association for Computing Machinery, New York, NY, USA, 203–210. <https://doi.org/10.1145/2049536.2049573>
- [65] Hernisa Kacorri. 2017. Teachable machines for accessibility. *SIGACCESS Access. Comput.* 119 (nov 2017), 10–18. <https://doi.org/10.1145/3167902.3167904>
- [66] Hernisa Kacorri, Kris M. Kitani, Jeffrey P. Bigham, and Chieko Asakawa. 2017. People with Visual Impairment Training Personal Object Recognizers: Feasibility and Challenges. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) (*CHI '17*). Association for Computing Machinery, New York, NY, USA, 5839–5849. <https://doi.org/10.1145/3025453.3025899>
- [67] Manisha Varma Kamarushi, Stacey L. Watson, Garreth W. Tigwell, and Roshan L. Peiris. 2022. OneButtonPIN: A Single Button Authentication Method for Blind or Low Vision Users to Improve Accessibility and Prevent Eavesdropping. *Proc. ACM Hum.-Comput. Interact.* 6, MHCI, Article 212 (sep 2022), 22 pages. <https://doi.org/10.1145/3546747>
- [68] Shaun K. Kane, Jeffrey P. Bigham, and Jacob O. Wobbrock. 2008. Slide rule: making mobile touch screens accessible to blind people using multi-touch interaction techniques. In *Proceedings of the 10th International ACM SIGACCESS Conference on Computers and Accessibility* (Halifax, Nova Scotia, Canada) (*Assets '08*). Association for Computing Machinery, New York, NY, USA, 73–80. <https://doi.org/10.1145/1414471.1414487>
- [69] Shaun K. Kane, Chandrika Jayant, Jacob O. Wobbrock, and Richard E. Ladner. 2009. Freedom to roam: a study of mobile device adoption and accessibility for people with visual and motor disabilities. In *Proceedings of the 11th International ACM SIGACCESS Conference on Computers and Accessibility* (Pittsburgh, Pennsylvania, USA) (*Assets '09*). Association for Computing Machinery, New York, NY, USA, 115–122. <https://doi.org/10.1145/1639642.1639663>
- [70] Shaun K. Kane, Jacob O. Wobbrock, and Richard E. Ladner. 2011. Usable gestures for blind people: understanding preference and performance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. Association for Computing Machinery, New York, NY, USA, 413–422. <https://doi.org/10.1145/1978942.1979001>
- [71] Cholmin Kang, Inhwa Yeom, Amirsaman Ashtari, Woontack Woo, and Junyong Noh. 2023. ARbility: re-inviting older wheelchair users to in-store shopping via wearable augmented reality. *Virtual Reality* 27, 3 (2023), 1919–1936.
- [72] Sunnie S. Y. Kim, Q. Vera Liao, Mihaela Vorvoreanu, Stephanie Ballard, and Jennifer Wortman Vaughan. 2024. 'I'm Not Sure, But...': Examining the Impact of Large Language Models' Uncertainty Expression on User Reliance and Trust. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency* (Rio de Janeiro, Brazil) (*FAccT '24*). Association for Computing Machinery, New York, NY, USA, 822–835. <https://doi.org/10.1145/3630106.3658941>
- [73] Patrick E. Lanigan, Aaron M. Paulos, Andrew W. Williams, Dan Rossi, and Priya Narasimhan. 2006. Trinetra: Assistive Technologies for Grocery Shopping for the Blind. In *2006 10th IEEE International Symposium on Wearable Computers*. 147–148. <https://doi.org/10.1109/ISWC.2006.286369>
- [74] Jaewook Lee, Yi-Hao Peng, Jaylin Herskovitz, and Anhong Guo. 2021. Image Explorer: Multi-Layered Touch Exploration to Make Images Accessible. In *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility* (Virtual Event, USA) (*ASSETS '21*). Association for Computing Machinery, New York, NY, USA, Article 69, 4 pages. <https://doi.org/10.1145/3441852.3476548>
- [75] Jaewook Lee, Andrew D. Tjahjaji, Jiho Kim, Junpu Yu, Minji Park, Jiawen Zhang, Jon E. Froehlich, Yapeng Tian, and Yuhang Zhao. 2024. CookAR: Affordance Augmentations in Wearable AR to Support Kitchen Tool Interactions for People with Low Vision. In *Proceedings of the Annual Symposium on User Interface Software and Technology (UIST '24)*.
- [76] Kyungyeon Lee, Sohyeon Park, and Uran Oh. 2021. Designing Product Descriptions for Supporting Independent Grocery Shopping of People with Visual Impairments. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) (*CHI EA '21*). Association for Computing Machinery, New York, NY, USA, Article 425, 6 pages. <https://doi.org/10.1145/3411763.3451806>
- [77] Franklin Mingzhe Li, Di Laura Chen, Mingming Fan, and Khai N. Truong. 2019. FMT: A Wearable Camera-Based Object Tracking Memory Aid for Older Adults. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 3, Article 95 (sep 2019), 25 pages. <https://doi.org/10.1145/3351253>
- [78] Franklin Mingzhe Li, Di Laura Chen, Mingming Fan, and Khai N. Truong. 2021. "I Choose Assistive Devices That Save My Face": A Study on Perceptions of Accessibility and Assistive Technology Use Conducted in China. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) (*CHI '21*). Association for Computing Machinery, New York, NY, USA, Article 372, 14 pages. <https://doi.org/10.1145/3411764.3445321>
- [79] Franklin Mingzhe Li, Jamie Dorst, Peter Cederberg, and Patrick Carrington. 2021. Non-Visual Cooking: Exploring Practices and Challenges of Meal Preparation by People with Visual Impairments. In *Proceedings of the 23rd International ACM*

- SIGACCESS Conference on Computers and Accessibility* (Virtual Event, USA) (*ASSETS '21*). Association for Computing Machinery, New York, NY, USA, Article 30, 11 pages. <https://doi.org/10.1145/3441852.3471215>
- [80] Franklin Mingzhe Li, Michael Xieyang Liu, Shaun K. Kane, and Patrick Carrington. 2024. A Contextual Inquiry of People with Vision Impairments in Cooking. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (*CHI '24*). Association for Computing Machinery, New York, NY, USA, Article 38, 14 pages. <https://doi.org/10.1145/3613904.3642233>
- [81] Franklin Mingzhe Li, Michael Xieyang Liu, Yang Zhang, and Patrick Carrington. 2022. Freedom to Choose: Understanding Input Modality Preferences of People with Upper-body Motor Impairments for Activities of Daily Living. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility* (Athens, Greece) (*ASSETS '22*). Association for Computing Machinery, New York, NY, USA, Article 39, 16 pages. <https://doi.org/10.1145/3517428.3544814>
- [82] Franklin Mingzhe Li, Lotus Zhang, Maryam Bandukda, Abigale Stangl, Kristen Shinohara, Leah Findlater, and Patrick Carrington. 2023. Understanding Visual Arts Experiences of Blind People. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (*CHI '23*). Association for Computing Machinery, New York, NY, USA, Article 60, 21 pages. <https://doi.org/10.1145/3544548.3580941>
- [83] Michal Luria, Rebecca Zheng, Bennett Huffman, Shuangni Huang, John Zimmerman, and Jodi Forlizzi. 2020. Social Boundaries for Personal Agents in the Interpersonal Space of the Home. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (*CHI '20*). Association for Computing Machinery, New York, NY, USA, 1–12. <https://doi.org/10.1145/3313831.3376311>
- [84] Kelly Mack, Emma McDonnell, Dhruv Jain, Lucy Lu Wang, Jon E. Froehlich, and Leah Findlater. 2021. What Do We Mean by “Accessibility Research”? A Literature Survey of Accessibility Papers in CHI and ASSETS from 1994 to 2019. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (*CHI '21*). Association for Computing Machinery, New York, NY, USA, Article 371, 18 pages. <https://doi.org/10.1145/3411764.3445412>
- [85] Tlameo Makati, Garreth W. Tigwell, and Kristen Shinohara. 2024. The Promise and Pitfalls of Web Accessibility Overlays for Blind and Low Vision Users. In *Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility* (St. John's, NL, Canada) (*ASSETS '24*). Association for Computing Machinery, New York, NY, USA, Article 38, 12 pages. <https://doi.org/10.1145/3663548.3675650>
- [86] Jennifer Mankoff, Gillian R. Hayes, and Devva Kasnitz. 2010. Disability studies as a source of critical inquiry for the field of assistive technology. In *Proceedings of the 12th International ACM SIGACCESS Conference on Computers and Accessibility* (Orlando, Florida, USA) (*ASSETS '10*). Association for Computing Machinery, New York, NY, USA, 3–10. <https://doi.org/10.1145/1878803.1878807>
- [87] Florian Mathis, Yu Sun, and Adrian Preussner. 2024. MRTranslate: Bridging Language Barriers in the Physical World Using a Mixed Reality Point-and-Translate System. In *Proceedings of the 2024 International Conference on Advanced Visual Interfaces* (Arenzano, Genoa, Italy) (*AVI '24*). Association for Computing Machinery, New York, NY, USA, Article 12, 9 pages. <https://doi.org/10.1145/3656650.3656652>
- [88] maxiAIDS products for independent living. 2024. *Color Identifier-Light Detector*. Retrieved Aug. 2024 from <https://www.maxiaids.com/product/color-identifier-light-detector#:~:text=The%20Colorino%20Color%20Identifier%20Light,and%20proximity%20by%20audio%20signal>
- [89] Emma J. McDonnell, Kelly Avery Mack, Kathrin Gerling, Katta Spiel, Cynthia L. Bennett, Robin N. Brewer, Rua Mae Williams, and Garreth W. Tigwell. 2023. Tackling the Lack of a Practical Guide in Disability-Centered Research. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility* (New York, NY, USA) (*ASSETS '23*). Association for Computing Machinery, New York, NY, USA, Article 106, 5 pages. <https://doi.org/10.1145/3597638.3615650>
- [90] Cecily Morrison, Edward Cutrell, Anupama Dhareshwar, Kevin Doherty, Anja Thieme, and Alex Taylor. 2017. Imagining Artificial Intelligence Applications with People with Visual Disabilities using Tactile Ideation. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility* (Baltimore, Maryland, USA) (*ASSETS '17*). Association for Computing Machinery, New York, NY, USA, 81–90. <https://doi.org/10.1145/3132525.3132530>
- [91] Cecily Morrison, Rita Faia Marques, Daniela Massiceti, Camilla Longden, Linda Wen, and Edward Cutrell. 2023. Understanding Personalized Accessibility through Teachable AI: Designing and Evaluating Find My Things for People who are Blind or Low Vision. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility* (New York, NY, USA) (*ASSETS '23*). Association for Computing Machinery, New York, NY, USA, Article 31, 12 pages. <https://doi.org/10.1145/3597638.3608395>
- [92] Martez E. Mott, Radu-Daniel Vatavu, Shaun K. Kane, and Jacob O. Wobbrock. 2016. Smart Touch: Improving Touch Accuracy for People with Motor Impairments with Template Matching. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) (*CHI '16*). Association for Computing Machinery, New York, NY, USA, 1934–1946. <https://doi.org/10.1145/2858036.2858390>
- [93] Vishnu Nair, Hanxiu 'Hazel' Zhu, and Brian A. Smith. 2023. ImageAssist: Tools for Enhancing Touchscreen-Based Image Exploration Systems for Blind and Low Vision Users. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (*CHI '23*). Association for Computing Machinery, New York, NY, USA, Article 76, 17 pages. <https://doi.org/10.1145/3544548.3581302>
- [94] Hugo Nicolau, João Guerreiro, Tiago Guerreiro, and Luís Carriço. 2013. UbiBraille: designing and evaluating a vibrotactile Braille-reading device. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility* (Bellevue, Washington) (*ASSETS '13*). Association for Computing Machinery, New York, NY, USA, Article 23, 8 pages. <https://doi.org/10.1145/2513383.2513437>
- [95] Royal National Institute of Blind People. 2024. *Talking Colour Detector DH289*. Retrieved Nov. 2024 from <https://shop.rnib.org.uk/talking-colour-detector>
- [96] Washington Group on Disability Statistics. 2022. *The Washington Group Short Set on Functioning (WG-SS)*. Retrieved January 26, 2024 from https://www.washingtongroupdisability.com/fileadmin/uploads/wg/Documents/Questions/Washington_Group_Questionnaire_1_-_WG_Short_Set_on_Functioning_June_2022_.pdf
- [97] Matthew Pateman, Daniel Harrison, Paul Marshall, and Marta E. Cecchinato. 2018. The Role of Aesthetics and Design: Wearables in Situ. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) (*CHI EA '18*). Association for Computing Machinery, New York, NY, USA, 1–6. <https://doi.org/10.1145/3170427.3188556>
- [98] Betsy Phillips and Hongxin Zhao. 1993. Predictors of Assistive Technology Abandonment. *Assistive Technology* 5, 1 (1993), 36–45. <https://doi.org/10.1080/10400435.1993.10132205> arXiv:<https://doi.org/10.1080/10400435.1993.10132205> PMID: 10171664
- [99] Halley Profita, Reem Albaghli, Leah Findlater, Paul Jaeger, and Shaun K. Kane. 2016. The AT Effect: How Disability Affects the Perceived Social Acceptability of Head-Mounted Display Use. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) (*CHI '16*). Association for Computing Machinery, New York, NY, USA, 4884–4895. <https://doi.org/10.1145/2858036.2858130>
- [100] Adil Rahman, Md Aashikur Rahman Azim, and Seongkook Heo. 2023. Take My Hand: Automated Hand-Based Spatial Guidance for the Visually Impaired. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (*CHI '23*). Association for Computing Machinery, New York, NY, USA, Article 544, 16 pages. <https://doi.org/10.1145/3544548.3581415>
- [101] Kyle Rector, Lauren Milne, Richard E. Ladner, Batya Friedman, and Julie A. Kientz. 2015. Exploring the Opportunities and Challenges with Exercise Technologies for People who are Blind or Low-Vision. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility* (Lisbon, Portugal) (*ASSETS '15*). Association for Computing Machinery,

- New York, NY, USA, 203–214. <https://doi.org/10.1145/2700648.2809846>
- [102] André Rodrigues, Kyle Montague, Hugo Nicolau, and Tiago Guerreiro. 2015. Getting Smartphones to Talkback: Understanding the Smartphone Adoption Process of Blind Users. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility* (Lisbon, Portugal) (*ASSETS '15*). Association for Computing Machinery, New York, NY, USA, 23–32. <https://doi.org/10.1145/2700648.2809842>
- [103] Daisuke Sato, Uran Oh, Kakuya Naito, Hironobu Takagi, Kris Kitani, and Chieko Asakawa. 2017. NavCog3: An Evaluation of a Smartphone-Based Blind Indoor Navigation Assistant with Semantic Features in a Large-Scale Environment. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility* (Baltimore, Maryland, USA) (*ASSETS '17*). Association for Computing Machinery, New York, NY, USA, 270–279. <https://doi.org/10.1145/3132525.3132535>
- [104] Matthias Schmidmaier, Jonathan Rupp, Darina Cvetanova, and Sven Mayer. 2024. Perceived Empathy of Technology Scale (PETS): Measuring Empathy of Systems Toward the User. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (*CHI '24*). Association for Computing Machinery, New York, NY, USA, Article 456, 18 pages. <https://doi.org/10.1145/3613904.3642035>
- [105] Ava Elizabeth Scott, Leon Reicherts, and Evropi Stefanidi. 2024. Doing CHI together: The benefits of a writing retreat for early-career HCI researchers. *Interactions* 31, 3 (2024), 8–9.
- [106] Roy Shilkrot, Jochen Huber, Connie Liu, Pattie Maes, and Suranga Chandima Nanayakkara. 2014. FingerReader: a wearable device to support text reading on the go. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems* (Toronto, Ontario, Canada) (*CHI EA '14*). Association for Computing Machinery, New York, NY, USA, 2359–2364. <https://doi.org/10.1145/2559206.2581220>
- [107] Kiyoungh Shin, Ryan McConville, Oussama Metatla, Minhye Chang, Chiyoungh Han, Junhaeng Lee, and Anne Roudaut. 2022. Outdoor Localization Using BLE RSSI and Accessible Pedestrian Signals for the Visually Impaired at Intersections. *Sensors* 22, 1 (2022). <https://doi.org/10.3390/s22010371>
- [108] Kristen Shinohara. 2012. A new approach for the design of assistive technologies: design for social acceptance. *SIGACCESS Access. Comput.* 102 (jan 2012), 45–48. <https://doi.org/10.1145/2140446.2140456>
- [109] Kristen Shinohara and Josh Tenenbergh. 2007. Observing Sara: a case study of a blind person's interactions with technology. In *Proceedings of the 9th International ACM SIGACCESS Conference on Computers and Accessibility* (Tempe, Arizona, USA) (*Assets '07*). Association for Computing Machinery, New York, NY, USA, 171–178. <https://doi.org/10.1145/1296843.1296873>
- [110] Kristen Shinohara and Jacob O. Wobbrock. 2011. In the shadow of misperception: assistive technology use and social interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) (*CHI '11*). Association for Computing Machinery, New York, NY, USA, 705–714. <https://doi.org/10.1145/1978942.1979044>
- [111] Kristen Shinohara and Jacob O. Wobbrock. 2016. Self-Conscious or Self-Confident? A Diary Study Conceptualizing the Social Accessibility of Assistive Technology. *ACM Trans. Access. Comput.* 8, 2, Article 5 (jan 2016), 31 pages. <https://doi.org/10.1145/2827857>
- [112] Kristen Shinohara, Jacob O. Wobbrock, and Wanda Pratt. 2018. Incorporating Social Factors in Accessible Design. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility* (Galway, Ireland) (*ASSETS '18*). Association for Computing Machinery, New York, NY, USA, 149–160. <https://doi.org/10.1145/3234695.3236346>
- [113] Hariharan Subramonyam, Steven M. Drucker, and Eytan Adar. 2019. Affinity Lens: Data-Assisted Affinity Diagramming with Augmented Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) (*CHI '19*). Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3290605.3300628>
- [114] Sarit Felicia Anais Szpiro, Shafeka Hashash, Yuhang Zhao, and Shiri Azenkot. 2016. How People with Low Vision Access Computing Devices: Understanding Challenges and Opportunities. In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility* (Reno, Nevada, USA) (*ASSETS '16*). Association for Computing Machinery, New York, NY, USA, 171–180. <https://doi.org/10.1145/2982142.2982168>
- [115] OrCam Learn Team. 2023. *Empowerment at Your Fingertips*. Retrieved January 30, 2024 from <https://www.orcam.com/>
- [116] Yong-Joon Thoo, Maximiliano Jeanneret Medina, Jon E. Froehlich, Nicolas Ruffieux, and Denis Lalanne. 2023. A Large-Scale Mixed-Methods Analysis of Blind and Low-vision Research in ACM and IEEE. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility* (*ASSETS '23*). Association for Computing Machinery, New York, NY, USA, Article 20, 20 pages. <https://doi.org/10.1145/3597638.3608412>
- [117] Garreth W. Tigwell, Benjamin M. Gorman, and Rachel Menzies. 2020. Emoji Accessibility for Visually Impaired People. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (*CHI '20*). Association for Computing Machinery, New York, NY, USA, 1–14. <https://doi.org/10.1145/3313831.3376267>
- [118] Daniel Trindade, André Rodrigues, Tiago Guerreiro, and Hugo Nicolau. 2018. Hybrid-Braille: Combining Physical and Gestural Interaction for Mobile Braille Input and Editing. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) (*CHI '18*). Association for Computing Machinery, New York, NY, USA, 1–12. <https://doi.org/10.1145/3173574.3173601>
- [119] Alexandra Vtyurina, Adam Fournery, Meredith Ringel Morris, Leah Findlater, and Ryen W. White. 2019. VERSE: Bridging Screen Readers and Voice Assistants for Enhanced Eyes-Free Web Search. In *Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility* (Pittsburgh, PA, USA) (*ASSETS '19*). Association for Computing Machinery, New York, NY, USA, 414–426. <https://doi.org/10.1145/3308561.3353773>
- [120] Ru Wang, Nihan Zhou, Tam Nguyen, Sanbrita Mondal, Bilge Mutlu, and Yuhang Zhao. 2023. Practices and Barriers of Cooking Training for Blind and Low Vision People. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility* (New York, NY, USA) (*ASSETS '23*). Association for Computing Machinery, New York, NY, USA, Article 57, 5 pages. <https://doi.org/10.1145/3597638.3614494>
- [121] Linda Yilin Wen, Cecily Morrison, Martin Grayson, Rita Faia Marques, Daniela Massiceti, Camilla Longden, and Edward Cutrell. 2024. Find My Things: Personalized Accessibility through Teachable AI for People who are Blind or Low Vision. In *Extended Abstracts of the 2024 CHI Conference on Human Factors in Computing Systems* (*CHI EA '24*). Association for Computing Machinery, New York, NY, USA, Article 403, 6 pages. <https://doi.org/10.1145/3613905.3648641>
- [122] Michele A. Williams, Erin Buehler, Amy Hurst, and Shaun K. Kane. 2015. What not to wearable: using participatory workshops to explore wearable device form factors for blind users. In *Proceedings of the 12th International Web for All Conference* (Florence, Italy) (*W4A '15*). Association for Computing Machinery, New York, NY, USA, Article 31, 4 pages. <https://doi.org/10.1145/2745555.2746664>
- [123] Jacob O. Wobbrock and Julie A. Kientz. 2016. Research Contributions in Human-Computer Interaction. *Interactions* 23, 3 (apr 2016), 38–44. <https://doi.org/10.1145/2907069>
- [124] Jingyi Xie, Rui Yu, He Zhang, Sooyeon Lee, Syed Masum Billah, and John M. Carroll. 2024. Emerging Practices for Large Multimodal Model (LMM) Assistance for People with Visual Impairments: Implications for Design. arXiv:2407.08882 [cs.HC] <https://arxiv.org/abs/2407.08882>
- [125] Miao Xiong, Zhiyuan Hu, Xinyang Lu, Yifei Li, Jie Fu, Junxian He, and Bryan Hooi. 2024. Can LLMs Express Their Uncertainty? An Empirical Evaluation of Confidence Elicitation in LLMs. arXiv:2306.13063 [cs.CL] <https://arxiv.org/abs/2306.13063>
- [126] Hui-Shyong Yeo, Erwin Wu, Daehwa Kim, Juyoung Lee, Hyung-il Kim, Seo Young Oh, Luna Takagi, Woontack Woo, Hideki Koike, and Aaron John Quigley. 2023. OmniSense: Exploring Novel Input Sensing and Interaction Techniques on Mobile Device with an Omni-Directional Camera. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (*CHI '23*). Association for Computing Machinery, New York, NY, USA, Article 530, 18 pages. <https://doi.org/10.1145/3544548.3580747>

- [127] Chun Yu, Xiaoying Wei, Shubh Vachher, Yue Qin, Chen Liang, Yueting Weng, Yizheng Gu, and Yuanchun Shi. 2019. Hand-See: Enabling Full Hand Interaction on Smartphone with Front Camera-based Stereo Vision. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) (CHI '19). Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3290605.3300935>
- [128] Lotus Zhang, Simon Sun, and Leah Findlater. 2023. Understanding Digital Content Creation Needs of Blind and Low Vision People. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '23)*. Association for Computing Machinery, New York, NY, USA, Article 8, 15 pages. <https://doi.org/10.1145/3597638.3608387>
- [129] Yuhang Zhao, Elizabeth Kupferstein, Doron Tal, and Shiri Azenkot. 2018. "It Looks Beautiful but Scary": How Low Vision People Navigate Stairs and Other Surface Level Changes. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility (Galway, Ireland) (ASSETS '18)*. Association for Computing Machinery, New York, NY, USA, 307–320. <https://doi.org/10.1145/3234695.3236359>
- [130] Yichun Zhao, Miguel A Nacenta, Mahadeo A. Sukhai, and Sowmya Somanath. 2024. TADA: Making Node-link Diagrams Accessible to Blind and Low-Vision People. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '24). Association for Computing Machinery, New York, NY, USA, Article 45, 20 pages. <https://doi.org/10.1145/3613904.3642222>
- [131] Yuhang Zhao, Sarit Szpiro, and Shiri Azenkot. 2015. 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 (Lisbon, Portugal) (ASSETS '15)*. Association for Computing Machinery, New York, NY, USA, 239–249. <https://doi.org/10.1145/2700648.2809865>
- [132] Annuska Zolyomi. 2024. Accessible Human-Computer Interaction + Inclusive Design. In *Teaching Accessible Computing*, Amy Oleson, Andrew J. Ko, and Richard Ladner (Eds.). <https://bookish.press/tac> Retrieved 13 Nov, 2024.

A CULTURAL PROBES: KICK-OFF MEETING, QUESTIONS AND SUPPORTING MATERIAL

The primary researcher met each participant at a location of their preference. The initial kick-off meeting served to introduce the participant to the cultural probes, collect demographic information, provide an overview of example everyday life scenarios (Figure 6), and distribute the video-recording equipment.

Table 3: Overview of participants' assistive technologies they mentioned during the kick-off meeting. Note that BLV people make use of a plethora of ATs and other assistance in their everyday life; therefore, this list is not meant to be exhaustive and representative of ATs used by BLV people in general. Instead, the list provides a glimpse into some of the systems the participants use to facilitate their day-to-day tasks.

Participant	Assistive Technologies
P1	Talking Watch; Rollator
P2	OrCam; Milestone; Screen Reader (PC + Phone)
P3	Color Detection Device; Seeing AI
P4	Milestone; MyWay Pro, Be My Eyes, Seeing AI; Talking Scale
P5/P5.a	Stationary OCR Device, Apps, Electronic Talking Pen
P6	Electronic Talking Pen; Seeing AI; Smart Shoes; OrCam
P7	Talking Scale
P8	Apps; Audio Description; Milestone; Talking Scale/Thermometer; OrCam
P9	Auditory Output on PC/Phone/Scale/Alarm Clock; OrCam; Milestone; Scanner

A.1 Questionnaire Kick-Off Meeting

We asked the following questions in the kick-off meeting to collect our participants' demographics.

- What is your gender? (Male, Female, Prefer not to say, Self-defined)
- How old are you?
- What type of living arrangement are you in? Please briefly describe where you live (Own home, Dormitory, etc.) and whether you live alone or with other people.
- Are there caregivers or family members who support you in your daily life? If yes, please briefly describe.
- Is your personal environment specifically adapted to your situation?
- Please describe your disability in 1-2 sentences.
- Since when do you have low vision/are you blind?
- Questions from the Washington Group Questionnaire [96]
- Could you please explain how your disability affects your everyday life? Please provide 2-3 examples of everyday situations where your disability plays a significant role.
- Do you have any restrictions on accessing certain places due to your disability? What are the consequences of this?
- Do you currently possess assistive technology that helps you in your daily life? This could be, for example, a specialized communication device, a special computer keyboard, etc.
- Have you used assistive technologies in the past that you no longer use? If yes, please briefly explain which technologies you used and why you stopped using them.
- Statement: "I am very familiar with technologies, such as smartphones, TVs, computers." (This statement is answered on a 5-point Likert scale (does not apply at all 1 - - - 5 fully applies)).
- Is there anything else you would like to tell us about your disability?

A.2 Supporting Material

We provided participants with additional information sheets about the use of the video camera. The materials were discussed in the kick-off meeting and sent to the participants via electronic mail.

- Step 1. Please attach the camera to the location discussed in the "Kick-Off Meeting" (e.g., using the shoulder strap).
- Step 2. The camera is equipped with four video presets – in this study, you use STANDARD (DEFAULT SETTING). This means you do not need to make any further adjustments to the camera.
- Step 3. Please press the shutter button to start recording, as introduced in the "Kick-Off Meeting". You can now record your everyday activities for as long as you want (up to 1.5 hours continuously due to the battery).
- Step 4. After recording, you can press the same button again. This will automatically stop the recording and power off the camera.



Figure 6: Activities of Daily Living, as visualized by Li et al. [81]. The figure includes daily activities such as bathing, dressing, grooming, oral care, toileting, transferring, moving around, eating, shopping, cooking, managing medication, using the phone, housework, laundry, driving, managing finance, leisure and other activities.

- Step 5. If the battery is running low, please recharge it. To do this, open the cover on the right side and plug the camera with the battery into the power outlet, as practised in the "Kick-Off Meeting". Ideally, the camera should be fully charged before each use. Alternatively, you can remove the battery and replace it with the second battery.
 - Step 1) Pull the release button on the right side of the camera towards you. The cover should then pop open.
 - Step 2) Remove the battery and replace it with the second battery.
- Step 6. If the memory space is running low, you should hear auditory feedback during the recording. Please replace the memory card (we provide you with 2 – 3 additional memory cards). To replace the memory card follow the following procedure, as practised in the "Kick-Off Meeting":
 - Hold down the release button of the battery door and slide the door open.
 - Remove the memory card and insert the second memory card.
- Please get in touch with us if you face any issues with the recording. We are happy to help.

B CULTURAL PROBES: CONCLUDING MEETING AND SEMI-STRUCTURED INTERVIEW QUESTIONS

The primary researcher scheduled a concluding session with each participant after the 1-week data collection phase. The concluding session served to create an anonymised backup of the cultural probes and held an additional semi-structured interview. The core questions are outlined below.

- Can you describe a typical day in your life and the specific challenges you experience due to your disability?
- What everyday activities present particular challenges for you due to your disability? [Discussion of filmed daily tasks]
- Are there situations or tasks where you urgently need help but currently have no suitable solutions?
- Can you share a positive experience with us where another person or a technology has made your day easier and more enjoyable regarding your disability?
- What improvements and changes in the public domain (e.g., infrastructure, transportation) would significantly improve your daily routine?
- What characteristics would a supportive technology need to have for you to use it?
- The 1 million [currency] question: Suppose you can wish for an ideal supportive technology that you want – without any limits. How would this technology look/-function?
- Overall, have there been any difficulties for you as part of this study?

C LIFEINSIGHT EVALUATION: SEMI-STRUCTURED INTERVIEWS AND MATERIAL

- **Topic 1: Usage Experience with LifeInsight**
 - **Example Question 1:** What contributed positively to your confidence in the prototype? What contributed negatively?
 - **Example Question 2:** How does the prototype compare to other assistive technologies you have used in the past? Are there specific features or functions that you prefer in this prototype over others?
- **Topic 2: Design and Interaction Space of LifeInsight**
 - **Example Question 1:** How did you find the interaction with the prototype using the single-click interaction?
 - **Example Question 2:** This system requires to have a view of your real-world surroundings using a camera. To achieve this, we attached a small camera to the collar of your clothing. On what other clothing items or body parts would you want to wear such a small camera?
- **Topic 3: Usage in Public Space and Social Acceptance**
 - **Example Question 1:** How would you feel if you would wear and interact with such a system in public spaces?
 - **Example Question 2:** What do you think others would perceive about you if you were to wear and use such a system in public spaces?
- **Topic 4: Motivation of Using LifeInsight in Daily Life**
 - **Example Question 1:** What would motivate you to use this prototype in your daily life?
 - **Example Question 2:** In which scenario would you prefer another person providing you with assistance? And in which scenario would you prefer to receive the assistance from such a system?

D CULTURAL PROBES: METADATA OF THE VIDEO RECORDINGS

Table 4 provides an overview of the video recordings that were collected during each participant session. Note that one session was scheduled for at least one week, with a concluding interview after the collection of the everyday life challenges.

E LIFEINSIGHT EVALUATION: NASA-TLX [59], SUS [35] AND PETS [104]

Table 5 provides a summary of the quantitative data collected during the empirical evaluation of LifeInsight.

F VISUAL QUERYING BEHAVIOURS

Table 6 shows participants' visual querying behaviours across all scenarios.

G LIFEINSIGHT EVALUATION: VISUAL SENSING BY PARTICIPANTS

Figure 7 provides an overview of example pictures taken by LifeInsight when used by the participants across the scenarios. All images were taken by BLV people during the empirical evaluation of LifeInsight.

Table 4: Overview of the metadata of the participant-driven video recordings. *# of Video Segments* represents the number of video segments recorded for each participant, *# of Used Video Segments* represents the distribution of the video segments that were used for each participant during the data processing. *Proportion* stands for the contribution to the overall video recordings dataset in percentages. *Duration* shows the average, min, and max duration of the video recordings in minutes (mm:ss). Videos that were too short to be meaningful or where we observed technical issues with the video-recording device are excluded. Videos that included multiple everyday life challenges were further split into video segments.

Participant	Primary Video Perspective(s)	# of Video Segments	# of Used Video Segments	Proportion	Duration [avg min max]		
P1	First- and Third Person	20	15	8.43%	05:34	01:30	11:48
P2	First Person	47	42	23.60%	03:01	00:13	09:33
P3	First Person	28	17	9.55%	01:05	00:33	02:13
P4	First Person	38	29	16.29%	06:29	00:29	11:56
P5	First- and Third Person	37	32	17.98%	03:01	00:16	09:02
P6	First Person	12	2	1.12%	01:01	00:47	01:15
P7	First Person	30	2	1.12%	00:52	00:46	00:59
P8	First Person	28	27	15.17%	07:37	00:30	14:07
P9	First Person	13	12	6.74%	01:07	00:12	02:21

Table 5: The table shows the NASA-TLX values for each dimension [59], the SUS [35], and the PETS [104].

Metric	LifeInsight
NASA-TLX [59]	
Mental Workload	M = 26.54 (SD = 21.45)
Physical Workload	M = 12.31 (SD = 14.23)
Temporal Workload	M = 17.31 (SD = 22.60)
Effort	M = 28.85 (SD = 16.97)
Performance	M = 24.23 (SD = 21.30)
Frustration	M = 26.15 (SD = 27.78)
Overall NASA-TLX	M = 22.56 (SD = 13.33)
SUS [35]	M = 74.81 (SD = 18.86)
PETS [104]	M = 43.85 (SD = 19.20)

Table 6: Overview of participants’ visual querying behaviours and the accuracy scores of LifeInsight in comparison to a human evaluation across the scenarios [52]. Values denote mean (standard deviation).

Visual Querying Behaviour	Localising	Understanding State	Devices & Interfaces	Comparing	Comprehending Text	Navigating
Incomplete Prompts	0.15 (0.55)	1.0 (1.08)	0.38 (0.77)	0.54 (1.39)	0.62 (1.12)	0.15 (0.38)
Generic Prompts	0.46 (0.97)	0.38 (0.96)	0.23 (0.60)	1.15 (2.19)	0.69 (1.03)	0.85 (1.07)
Object-centred Prompts	0.54 (0.88)	0.85 (2.23)	0.92 (1.26)	2.62 (3.40)	1.54 (2.18)	0.77 (1.36)
Targeted Object-centred Prompts	1.46 (1.20)	1.62 (1.12)	2.62 (2.96)	0.31 (0.63)	2.15 (2.91)	1.62 (1.50)
Overall Number of Prompts	2.62 (1.19)	3.85 (3.18)	4.15 (3.31)	4.62 (3.43)	5.00 (4.49)	3.38 (2.43)
Accuracy Scores	2.51 (0.68)	2.92 (0.20)	3.0 (0.0)	2.85 (0.40)	2.49 (0.77)	2.03 (0.78)

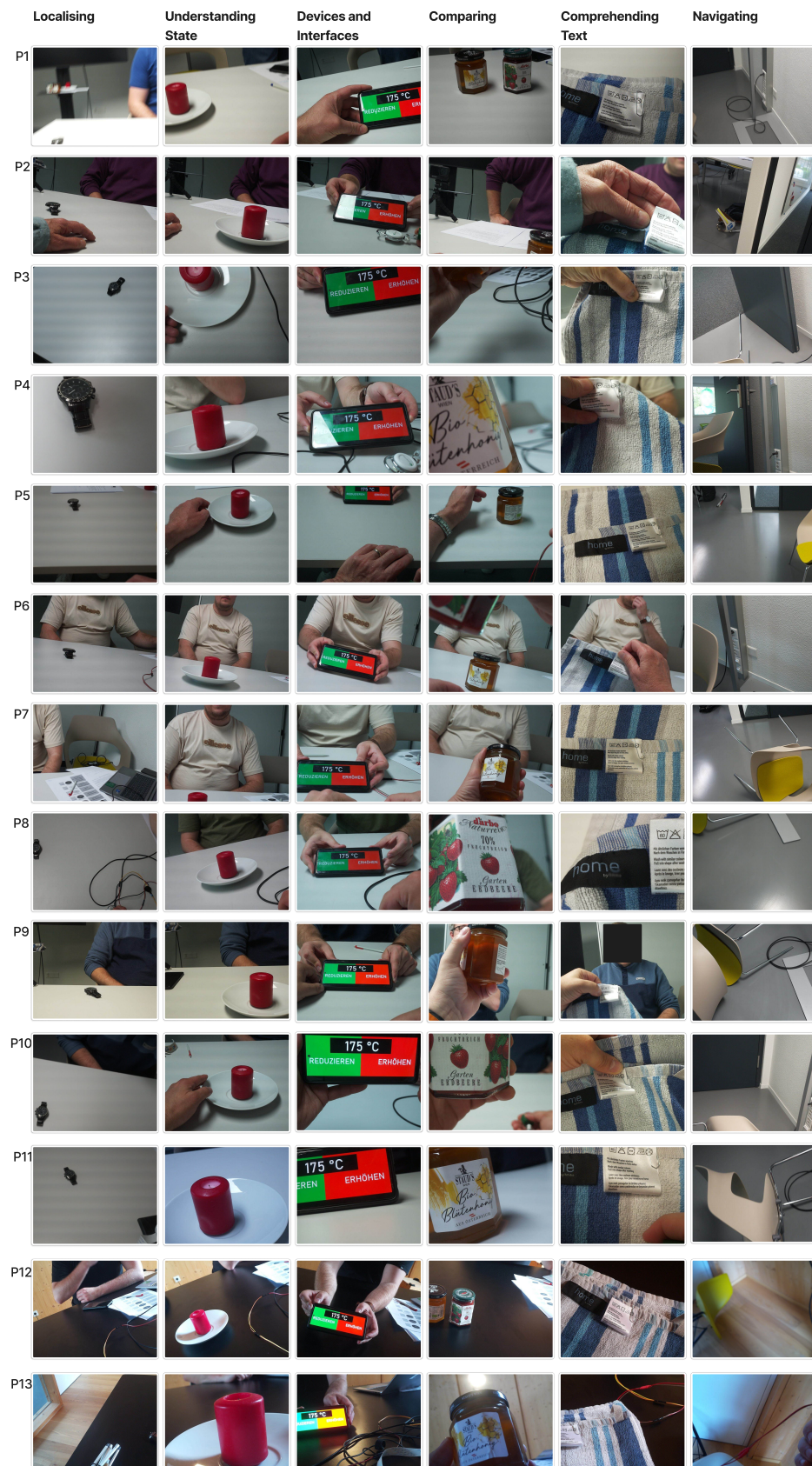


Figure 7: Images taken by BLV people when using LifeInsight across the six scenarios. Each row represents one participant.