# "Trying to Piece It Together": Exploring Accessible Error Detection in Emerging Privacy Techniques With Blind People


Rahaf Alharbi*
School of Information
University of Michigan
Ann Arbor, Michigan, USA
Pratt Institute
New York, USA
ralhar15@pratt.edu

Angela D Cheong
School of Information
University of Michigan
Ann Arbor, Michigan, USA
acheong@umich.edu

Jaylin Herskovitz
Computer Science and Engineering
University of Michigan
Ann Arbor, Michigan, USA
jayhersk@umich.edu

Robin N. Brewer
School of Information
University of Michigan
Ann Arbor, Michigan, USA
rnbrew@umich.edu

Sarita Schoenebeck
School of Information
University of Michigan
Ann Arbor, Michigan, USA
yardi@umich.edu



## Abstract

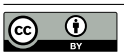
Blind people use visual assistance technologies (VAT) to access visual information, yet VAT can expose blind people to privacy risks. Prior HCI research has studied and built AI-enabled obfuscation techniques to detect and remove private content. However, blind people cannot easily spot errors in obfuscation tools. Our paper explores how assessment descriptors, brief visual attributes of objects, may enable blind people to find errors. By conducting interviews and focus groups with blind participants, we found that certain assessment descriptors (color, dimensions, distance) are inadequate to support blind people. Instead, participants discussed assessment descriptors that better reflect their sensemaking process, such as describing multiple objects in a particular space. Expanding the scope of accessible verification beyond assessment descriptors, participants called for greater transparency on how AI-enabled privacy techniques are developed and emphasized the need to co-create training materials on using AI-enabled privacy techniques. Building from our findings and disability studies scholarship, our paper examines how sighted bias could produce assessment descriptors that neglect the needs of blind people and analyzes how participants' preferred assessment descriptors contrast with existing standards of visual description. Lastly, we offer design directions to push for greater transparency in VAT and obfuscation tools.



## CCS Concepts

• **Human-centered computing** → **Empirical studies in accessibility**.


*This work was conducted while Alharbi was a doctoral student at the University of Michigan.





*ASSETS '25, Denver, CO, USA*



ACM ISBN 979-8-4007-0676-9/25/10
https://doi.org/10.1145/3663547.3746376

## Keywords

Visual Assistance Technologies, Privacy, Obfuscation, Artificial Intelligence, AI-enabled Privacy Techniques, Accessible Verification, Transparency

**ACM Reference Format:**
Rahaf Alharbi, Angela D Cheong, Jaylin Herskovitz, Robin N. Brewer, and Sarita Schoenebeck. 2025. "Trying to Piece It Together": Exploring Accessible Error Detection in Emerging Privacy Techniques With Blind People. In *The 27th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '25), October 26–29, 2025, Denver, CO, USA.* ACM, New York, NY, USA, 20 pages. https://doi.org/10.1145/3663547.3746376


## 1 Introduction

Blind people use visual assistance technologies (VAT) to gain access in their everyday lives, from reading mail to picking out fashionable outfits [26, 31, 56, 62]. Despite the benefits of VAT, they often include privacy implications. The possibility of inadvertently capturing sensitive information in the background of a photo (e.g., personal mail, pregnancy tests, or family photos) and sending that photo to an online AI or human-assistance service is a real and pressing risk [4, 34, 52, 99]. Nevertheless, blind people may feel compelled to tolerate these privacy risks in the absence of other accessible options [6, 99]. To readdress and mitigate risks, researchers have built *obfuscation technologies*: artificial intelligence[1] (AI)-enabled techniques that automatically detect and remove private content by applying filtering techniques such as blurring [6, 52, 97, 118, 119].

However, like all AI technologies, obfuscation techniques are imperfect and may contain errors. For instance, they may misrecognize objects that should be obscured (e.g., labeling a condom packet as a toy) so that they are mistakenly revealed [118]. Furthermore, these errors are particularly difficult for blind users to identify. While blind people are enthusiastic about controlling their visual data with obfuscation, they worry about potential obfuscation errors (e.g., obfuscation systems would wrongly confirm the

[1]In this paper, we use AI as an umbrella term. Obfuscation techniques typically include image processing methods [9], various deep learning models [68, 69, 116], and multimodal large language models [84].

detection and removal of private content), which are challenging to find non-visually [6, 97, 118]. Accordingly, scholars have advocated for developing non-visual transparency to empower blind people in detecting obfuscation errors [6, 97].

**This paper explores whether and how *assessment descriptors* could support blind people in detecting obfuscation errors.** Specifically, assessment descriptors are key visual attributes of obfuscated or spotlighted content that provide blind users with additional information to confirm or reject obfuscation. For example, if a blind user wanted to understand whether a photo on their wall is obfuscated, assessment descriptors may describe visual cues like 'a photo with a brown wooden frame.' Assessment descriptors have been used in other accessibility contexts to aid blind people in building their understanding of visual content. For instance, Hong et al. [2021] explored using descriptors, such as image quality and object size, for blind people to inspect AI training images [58]. In contrast to AI confidence scores that are difficult to interpret [6, 7, 97], assessment descriptors may offer richer means for blind users to engage with obfuscation errors. From a technical standpoint, assessment descriptors could be especially valuable because they can be generated on-device, minimizing potential security risks associated with off-device or remote processing [70].

We conducted a two-part qualitative study to explore the promise and limitations of assessment descriptors. First, we interviewed 26 blind participants to introduce and elicit perspectives on obfuscation techniques more broadly. Next, we invited participants to focus groups to discuss how assessment descriptors may support (or hinder) their process of detecting obfuscation errors. To ground our discussion, we presented pre-recorded audio probes that verbalized user interactions with obfuscation techniques. Specifically, we examined assessment descriptors such as an object's color, dimensions, and distance from the user, and aimed to prompt reflections on potential benefits and harms. In total, we facilitated seven focus groups with 16 participants.

Our findings suggest that assessment descriptors such as color, distance, and location are misaligned with how blind people identify objects and may not contribute to finding obfuscation errors. Alternatively, participants wanted assessment descriptors to name objects and include relevant materials such as text. They also desired information about surrounding objects, referencing assessment descriptors with their familiarity with the space to validate obfuscation. Lastly, participants identified structural and technical requirements needed to make assessment descriptors work. They discussed the importance of customizing information depth and system-level transparency to understand how obfuscation operates. Collectively, our analysis unpacks the technical and organizational facets of emerging privacy techniques in visual assistance technologies.

We make three primary contributions. The first is a detailed account of blind people's perspectives on accessible error detection in emerging AI-enabled privacy techniques. While past HCI and accessibility works studied and built obfuscation tools for blind people [6, 52, 97, 116], strategies to empower blind people in catching errors are underexplored. Our findings corroborate and build from the fields of accessible privacy techniques [6, 97, 110] and non-visual inspection of AI output [7, 58, 60, 74]. Second, we categorize how sighted bias negatively influences the design of assessment descriptors, draw parallels to existing scholarship on visual description, and offer directions to challenge harmful design choices when developing assessment descriptors. Additionally, we articulate directions to expand the scope of accessible error detection beyond assessment descriptors. Third, we derive general implications for accessible privacy techniques, highlighting the need to develop training materials and support infrastructure with blind communities.

## 2 Related Work

### 2.1 Visual Assistance Technologies (VAT)

Blind people use visual assistance technologies (VAT) to gain visual access. Broadly, VAT are mobile applications that take videos and images as input, and output verbal or haptic descriptions [15, 89, 94]. More recently, VAT started including desktop applications [14, 46] and wearable glasses [43, 63]. Typically, VAT are categorized based on who or what mediates visual information - humans or artificial intelligence (AI). With human-enabled VAT, remote volunteers (e.g., Be My Eyes [15]), trained agents (e.g., Aira [3]), or crowdworkers [25] provide visual access with blind people. AI-enabled VAT (e.g., Seeing AI [94], Be My AI [89], and Envision AI [42]) uses computer vision and large language models to describe visual content. HCI and accessibility research has sought to understand how blind people use VAT [7, 12, 26, 45, 56]. By reviewing over 1000 images taken by blind people, Brady et al. [2013] outlined three key visual tasks: (1) identifying (i.e., naming) objects, (2) reading textual content like mail, and (3) describing visual properties such as the color of a t-shirt [26]. These general cases reveal the differences between AI- and human-enabled VAT. AI-enabled VAT are typically used for "objective" tasks like reading [88], whereas human-enabled VAT are often used for complex tasks such as fashion advice [31]. While AI-enabled technologies are perceived to be more convenient, they are prone to errors [2, 7] and are sometimes disconnected from the real-world needs of blind people [7, 47, 56]. In a literature review of AI-enabled assistive technologies for blind people, Gamage et al. [2023] demonstrated that 82% of studies did not involve blind people [47]. Accordingly, they report a difference between the desires of blind people and the AI systems researchers created.

Accessibility research has begun taking a community-centered approach to building VAT *with* blind people [47, 51, 56]. Through an interview and diary study, Herskovitz et al. [2023] introduced opportunities to support blind people in customizing VAT [56]. Morrison et al. [2023] drew from the principles of citizen science [95] to co-design "Find My Things" (now a feature in Seeing AI [94]) with blind communities [86]. We add to this growing body of literature by conducting an in-depth qualitative study with blind people on their perspectives of future VAT features, revealing insights on how certain designs may complement (or clash) with their everyday use.

### 2.2 The Privacy Implications of VAT & Proposed Safeguards

Like any camera-based technology, VAT pose privacy concerns. Past research categorized and studied privacy risks associated with VAT use [4, 52, 98, 99]. Gurari et al. [2019] found that 10% of images submitted to VizWiz (a type of VAT) included private content, such as pregnancy tests, prescription medication, and people

[52]. Alarmingly, the majority of these privacy leaks were in the background, indicating that they might be accidental. Stangl et al. [2020] evaluated blind people's concerns during unknown and known disclosures of private content in human and AI-enabled VAT and found participants reported more significant concerns with unknown disclosures [99]. Akter et al. [2020] studied blind people's privacy concerns in human-enabled VAT, demonstrating concerns over value judgments and identity theft [4]. Collectively, these studies call to develop technologies to mitigate privacy risks.

Specifically, prior research suggested and built obfuscation tools, an AI-enabled technique that detects and filters (e.g., blur or mask) private content [52, 118, 119]. Obfuscation methods can be applied to specific regions of interest (partial obfuscation) [8, 27, 55] or to the full visual content (total obfuscation) [39]. Notably, there a few empirical studies that examine blind people's perceptions and experiences with obfuscation. Zhang et al. [2024] built an obfuscation prototype and evaluated its utility with blind people for content creation uses [118]. Their findings revealed that blind people were enthusiastic about the potential to preserve visual privacy before posting images to social media, yet they experienced cognitive overload. While some of these findings may apply to VAT, it is important to note that image sharing on social media is a different context than VAT. Social media images are static and often require attention to visual aesthetics [120]. VAT includes both images and videos, primarily for visual information seeking. More related to our study, Alharbi et al. [2022] investigated blind people's perspectives on using obfuscation to address privacy risks in VAT [6]. They found that blind people wanted to enact more control over obfuscation by choosing when to apply obfuscation and determining obfuscation content. Similarly, Stangl et al. [2023] found that blind people emphasized the need to design for effective consent and dismissal in obfuscation systems [97].

We add to prior research by investigating two types of obfuscation intended to provide control and choice: (1) focus mode, in which a specific object is spotlighted, and all other elements are obfuscated or hidden, and (2) background mode, in which a specific object is obfuscated, and all other elements are preserved. While past studies reported that blind people imagined benefiting from a feature such as focus mode [6, 97], they did not directly probe for focus mode. Furthermore, based on participants' desires to learn how to use emerging AI-enabled privacy tools, our analysis offers insights on developing training materials for obfuscation.

## 2.3 How Users Find & Detect Errors

Prior obfuscation research found that blind people were concerned about errors, especially in high-risk and complex use cases such as navigation [6]. Similarly, Stangl et al. [2023] reported that blind people worried about false positive and false negative errors with obfuscation in VAT [97]. For instance, past work that evaluated an obfuscation prototype with blind people demonstrated that obfuscation misrecognition was prevalent, and blind participants could not reliably validate obfuscation results [118]. Responding to these concerns, our study aims to support blind people in detecting obfuscation errors.

Generally, there is a wealth of HCI and AI scholarship dedicated to understanding how *sighted* users recognize and resolve errors (e.g., [32, 44, 61, 67, 75]). Goloujeh et al. [2024] found that users deployed various evaluation strategies with AI-generated images, such as visually inspecting the aesthetic of different outputs and noticing errors or preferences [75]. In studying an AI-enabled bird identification application, Kim et al. [2023] found that users validated the recognition quality by visually comparing the output with images online [67]. More related to our work, past research found that sighted users trusted obfuscation systems because they could visually confirm that private content is obfuscated [9]. Taken together, these studies signal an overemphasis on visual sensibilities in AI error detection approaches.

A few past studies focused on understanding the types of AI errors blind people encounter and how they verify AI [1, 7, 74]. Abdolrahmani et al. [2017] categorized blind people's error acceptance in AI tools for navigation, indicating that blind people worried about stigmatizing errors such as misidentifying gender on bathroom signs [1]. Alharbi et al. [2024] described how blind people detect errors in AI-enabled VAT, reporting strategies such as experimenting in low-risk settings and cross-referencing with different applications [7]. Adnin & Das [2024] found that blind people detected generative AI errors by comparing them with their previous knowledge or requesting that AI systems "prove" its answer [2]. In high-stakes cases (e.g., financial information), blind people would take extra steps to ensure accuracy, such as including sighted people [2, 6]. Overall, previous research highlighted that blind people often evaluate the risks of incorrect AI input and decide whether it is worth investing time in verifying potential errors. However, *how* to support blind people in finding AI errors remains unexplored. In the context of automatic alternative text, MacLeod et al. [2017] found that incorporating confidence scores could help blind people critically assess accuracy [74]. Nevertheless, blind participants in prior work on obfuscation noted that confidence ratings might be challenging to interpret meaningfully and could be misleading (e.g., conveying a high confidence rating despite false predictions) [6]. One closely related study investigated the use of additional descriptors to support blind people in training AI. Specifically, Hong et al. [2022] explored adding information on object size, image quality, and location within the image, showing how these descriptors, though occasionally inaccurate, helped guide blind users to improve their training images [59].

Our study investigates blind people's perspectives on error detection in obfuscation techniques. We examine the potential and limitations of *assessment descriptors*, visual attributes that describe elements such as color, size, and location of an object before obfuscation. Our analysis demonstrates how the design of some assessment descriptors may be influenced by sighted bias, and details participants' preferred assessment descriptors. Furthermore, we outline ways to support accessible error detection beyond assessment descriptors.

# 3 Method

We took a qualitative approach to investigate how blind people imagined making sense of emerging privacy techniques in VAT. First, we conducted interviews to introduce participants to two possible obfuscation techniques (focus mode and background) and elicit their broader reactions on the potential and drawbacks of

obfuscation. Next, to dive deeper into accessible error detection, we ran a series of focus groups on how assessment descriptors may support blind people in finding obfuscation errors. This section details our methodological procedure, recruitment strategy, and data analysis approach.

## 3.1 Procedure

*3.1.1 Initial Interviews.* We conducted interviews as part of a larger study on how blind people interpret and experience AI errors more broadly. This paper analyzes the last segment of our prior interviews [7], which lasted approximately 20-30 minutes and aimed to introduce obfuscation to participants and gather their perspectives on finding errors in emerging AI-enabled privacy approaches. Particularly, we defined (1) **focus mode**, which involves spotlighting a specific object while obfuscating or hiding all other elements, and (2) **background mode**, which entails obfuscating a private object while preserving the rest of the image or video. While past work primarily introduced AI-enabled privacy techniques as tools to remove one private content by applying filters (i.e., background mode), we added focus mode because blind participants in previous research envisioned benefiting from an ability to highlight one identified object while obfuscating remaining content [6, 97, 118]. Similar to prior work [6, 97], we tried to use plain language descriptions, avoiding technical jargon like "artificial intelligence", "machine learning" or even "automatic" since these might introduce biases. Appendix A includes a script of how we described focus and background modes to participants. After each instance of introducing focus mode and background mode, we asked participants about their thoughts, potential benefits/harms to themselves or the community, and scenarios where they would (not) prefer to use focus or background mode. We also inquired about accuracy and how they imagined assessing the credibility of focus or background mode. During the interviews, we intentionally did not introduce assessment descriptors. This decision aimed to reduce cognitive load since participants were already asked to consider two emerging techniques and to prevent influencing participants' views on how accessible error detection should be designed. Appendix B provides an overview of the questions we have asked during the interviews.

*3.1.2 Focus Groups.* We designed 60-90 minute semi-structured focus groups to further understand participants' perspectives on two emerging privacy techniques in VAT and the role of assessment descriptors in aiding or hindering error detection. In what follows, we will elaborate on the components of our focus group.

**Group Norms:** Before scheduling focus groups, we explained to participants that privacy expectations differ from interviews. While the research team must maintain privacy practices, we informed participants that other participants in the focus groups are not formally obligated to preserve their privacy. We indicated some steps participants can take to safeguard their privacy (e.g., changing Zoom names to pseudonyms). We also attached our focus group protocol so participants can read the questions ahead of time and anticipate privacy-persevering responses. Finally, we reminded participants that they could skip questions without explaining why. During the focus groups, we asked participants to refrain from sharing others' perspectives outside of this focus group, engage with each other's responses, and minimize cross-talk. We emphasized that the goal of our focus groups was not to establish collective agreements; embracing differences and respectful disagreements are welcomed. Additionally, we asked participants if they had any norms they would like us to follow. Our group norms were adapted from the Program on Intergroup Relations at the University of Michigan [102].

**Fictional Audio Probes:** To explore assessment descriptors, we designed fictional pre-recorded audio probes based on obfuscation use cases from prior work [4, 6, 97]. For instance, we chose scenarios in home settings to elicit potential feelings of impression management [4, 6]. Figure 1 captures the main elements of our probes (appendix D includes the scripts used for each probe). Our goal with these illustrative examples was to inspire conversations about accessible error detection in obfuscation. With that in mind, we presented various assessment descriptors: the "private" object's color, size, and location. Inspired by accessibility features that use metrics like distance to support visual access [10, 58, 101], we chose measurements as a potential assessment descriptor. Additionally, we included color as an assessment descriptor because AI systems often misrecognize color, prompting participants to think more critically about possible errors [47, 58, 105]. Overall, we emphasized that these hypothetical probes may not align exactly with participants' experiences, and we informed participants that the audio probes were intentionally designed to be short to enable us to build a narrative together. While the audio probes helped in accessibly imagining obfuscation techniques, later parts of this paper elaborate on key limitations that could have been resolved by conducting formative studies before focus groups. We encourage future research to construct audio probes in collaboration with blind communities.

**Followup Questions:** After playing each audio probe, we asked participants several questions to capture their perceptions on the presented assessment descriptor(s) and understand how (if at all) the presented assessment descriptors might be applicable to their everyday use of VAT. We also asked participants to comment on concerns, probing for specific examples of when and how assessment descriptors may hinder accessibility. Next, we asked participants what questions they would have before using focus and background modes and how they would imagine their roles in improving these systems moving forward (if any). Finally, we concluded with space for participants to add thoughts and considerations that the focus group may have missed and ask the lead researcher any questions. Appendix C includes an overview of the questions we asked during our focus groups.

## 3.2 Recruitment and Participation

After receiving ethics review approval for our study, we collaborated with the National Federation of the Blind (NFB) to recruit participants. We designed a short recruitment survey to confirm eligibility (i.e., participants are over 18 years old, reside in the United States, and use VAT) and inquire about the types of VAT respondents' use. To ensure diversity in our sample, we included optional demographic questions around race, gender, age, level of visual disability, and when they acquired their visual disability. As a token of appreciation for their expertise and time, participants were

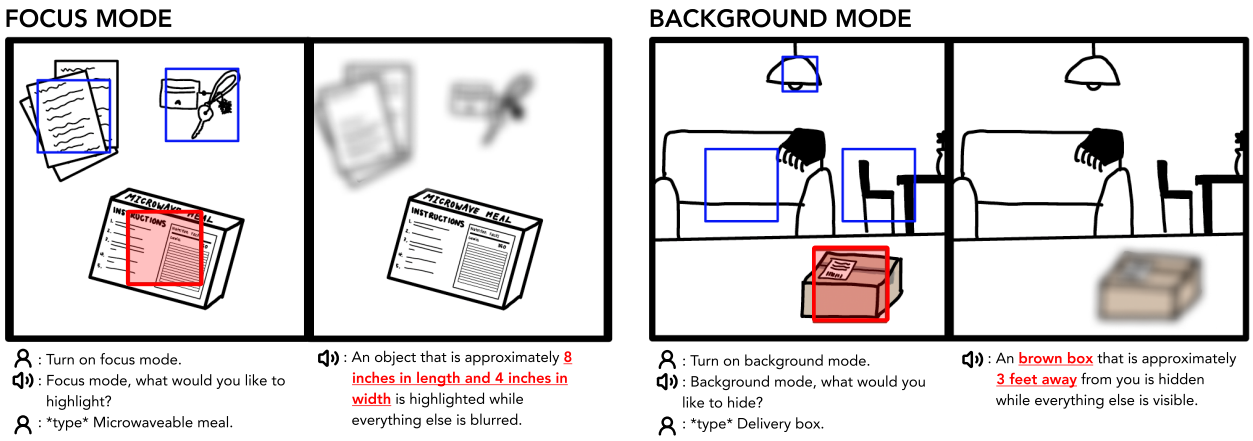



**Figure 1: Illustration of hypothetical scenarios captured in audio probes. In focus mode, we used the dimensions of the object as an assessment descriptor. In background mode, we explored the location and color of the object as an assessment descriptor.**

**Table 1: Breakdown of focus groups (FG), participant IDs, and visual assistance technologies (VAT) use**

| FG | Participant(s) ID | Visual Assistance Technologies Used |
|---|---|---|
| FG1 | P20 | Be My Eyes, Be My AI, Seeing AI, OneStep Reader |
| FG2 | P3, P12 | Aira, Seeing AI, Be My Eyes, OneStep Reader, BeSepcular, Envision AI |
| FG3 | P4, P10, P14 | Be My Eyes, Be My AI, Seeing AI, TapTapSee, Envision AI, OneStep Reader |
| FG4 | P2, P8, P22 | Be My Eyes, Seeing AI, Google Lookout, TapTapSee |
| FG5 | P7, P16 | Aira, Seeing AI, Be My Eyes, Be My AI, TapTapSee, Envision AI |
| FG6 | P23, P24 | Aira, Be My Eyes, Be My AI, Seeing AI, Envision AI, OneStep Reader |
| FG7 | P6, P9, P19 | Aira, Be My Eyes, Be My AI, Seeing AI, One Step Reader, TapTapSee |

compensated with $35 (USD) for the interviews and $55 (USD) for the focus groups.

We conducted interviews from September to October 2023. Our interviews included 26 participants. All participants were daily or weekly users of VAT, particularly Seeing AI, Be My Eyes, Be My AI, Aira, TapTapSee, and Envision AI. To preserve anonymity, we report an aggregated demographic of the 26 interview participants. In terms of visual disability, the majority of participants were totally blind (n=25), and one participant had low vision. Most (n=15) were born with a visual disability; the remaining (n=11) acquired their disability later in life during childhood or adulthood. We inquired about their age through a multiple-choice question with age range (e.g., 18-24, 25-34, etc). The weighted average of participants who reported their age (n=25) was forty-one years old. As for gender, most of our participants (n=15) are women, and some (n=11) are men. Twenty-three participants opted to report racial or ethnic identity. Our sample included participants who are White (n=10), Hispanic (n=3), Latinx (n=2), Asian (n=4), Middle Eastern (n=2), and mixed race (n=2).

After interviews, we invited participants who indicated interest in the focus groups. We scheduled and conducted seven focus groups with 16 participants in November 2023. Table 1 provides a breakdown of participants. Those who participated in the focus groups were all totally blind. Ten do not have light perception, and six have light perception. Nine acquired their visual disability at birth, and seven later in life during childhood or adulthood. In terms of gender, nine are women, and seven are men. Our focus group sample included participants who are White (n=6), Asian (n=3), Middle Eastern (n=2), Hispanic (n=1), Latinx (n=1), mixed race (n=1), and not reported (n=2).

## 3.3 Data Analysis & Positionality

We took a reflexive thematic approach [28, 29] to analyze the interview and focus group transcripts. The first author spearheaded the analysis process with the help of the second author. We first began familiarizing ourselves with the data, spending over two months (re)reading each transcript, writing memos, and discussing patterns. Then, we open-coded all the transcripts. The first author read through all the quotes associated with each code. We then organized quotes in a Miro[2] board based on emerging patterns such as potential use cases of privacy features, reactions to assessment descriptors, suggestions for assessment descriptors, error handling, and transparency. Figure 2 represents an overview of our early analysis on Miro. In grouping the quotes from the focus groups, we aimed to create a cohesive story through meaning-based interpretation rather than providing a topic summary of each focus group. In line with reflexive thematic analysis sensibilities [30], we sought to portray each participant's quote in one focus group through discussions across our dataset, not just in the bounds of their focus group.

[2]Miro: https://miro.com/

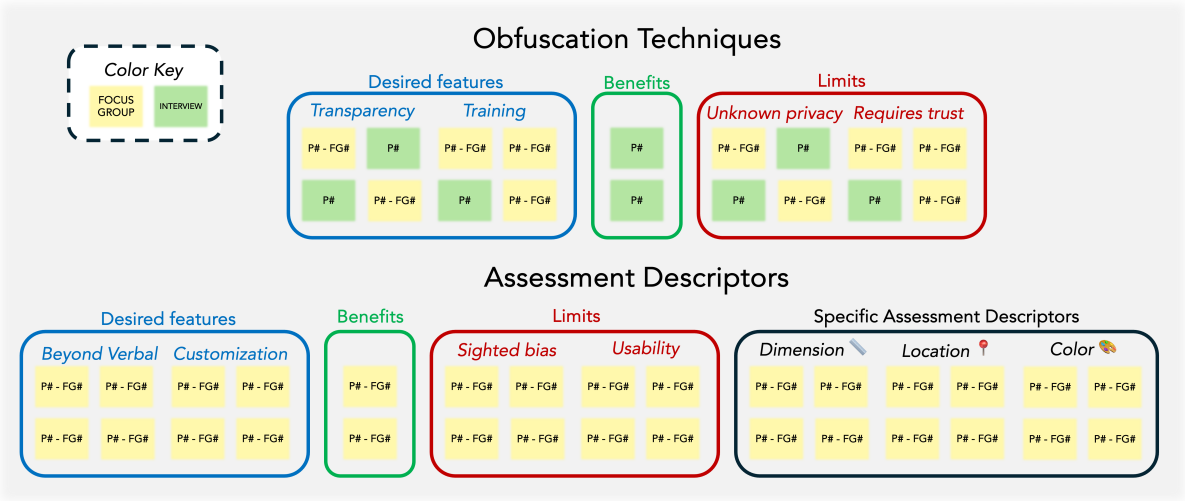



**Figure 2: Recreated skeleton of our preliminary data analysis on Miro, capturing some early patterns and codes. We initially mapped participants' quotes on obfuscation techniques broadly, then we organized participants' perspectives on assessment descriptors.**

After discussions with co-authors and four additional iterations of refining patterns, the first author structured the corpus into the themes that now serve as findings.

Reflexive thematic analysis fosters recognizing positionality, attending to how the researchers' views (in)directly shaped research questions and data interpretation. Members of this research team are sighted accessibility researchers who are grounded in and often draw from disability studies and disability justice activism [21, 76, 100]. We acknowledge the limits of our knowledge as sighted people. We took the following steps to work within power asymmetries [19] and practice reflexivity. The first author, who led interviews and focus groups, began sessions affirming participants' expertise and emphasizing that there are no right or wrong answers. We believe this cultivated a space where participants feel comfortable providing honest responses and even challenging how the interviewer framed some questions. During the interviews and focus groups, the lead author performed member checking to assess their interpretation and offer participants an opportunity to elaborate. Before concluding interviews and focus groups, we incorporated a closing question [79] to allow participants to share topics we may have missed or ask the lead researcher any questions.

## 4 Findings

In this section, we report findings from interviews[3] and focus groups. Confirming prior work, participants saw high value in obfuscation features and expressed concern about errors. However, we found that most participants did not find assessment descriptors such as color, dimension, and location useful. Participants discussed how vague and highly visual assessment descriptors are not aligned with their sensemaking processes. Alternatively, participants suggested assessment descriptors that directly name objects, incorporate surrounding objects and support haptic and auditory feedback. Beyond assessment descriptors, participants emphasized the importance of explaining to users how AI-enabled privacy techniques work and their limitations. They also outlined opportunities for tutorials on how to use AI-enabled technologies. Our findings capture technical and organizational considerations for building AI-enabled privacy techniques aligned with blind communities.

### 4.1 Critiques and Anticipated Challenges of Assessment Descriptors

In the audio probes, we highlighted assessment descriptors that were realistic and used in previous related contexts to provoke discussion. However, the majority of participants did not find these particular assessment descriptors helpful, citing accuracy concerns, lack of clarity, added cognitive load when using human-enabled VAT, and the necessity of prerequisite visual knowledge to confirm assessment descriptors.

*4.1.1 Questioning the Accuracy of Assessment Descriptors.* For assessment descriptors to work, they need to be accurate. However, participants hesitated to trust assessment descriptors and, thus, obfuscation tools. P12 emphasized, *"I would still have a trust issue. Is it really blurring out all the information? I'm not sure how you could assure me without being able to see the screen and see it blurred out."* P24 added, *"you don't have direct access to the information. It's the computer, the app, whatever is interpreting it for you."* Participants' concerns about the accuracy of assessment descriptors stem from prior experience using AI systems. For instance, P8 reflected on how AI-enabled VAT struggled to describe the basic shapes in their dog's raincoat, asserting that AI systems are *"not humble"* and will confidently report false information. Additionally, participants raised concerns about how the object's location and camera aiming would shape accuracy. P6 explained, *"because some people hold it at*

[3]**Note for readers:** Participant IDs marked with a double dagger (‡) indicate that the related quote was shared during the interview portion of the study.

*an angle, some people place it right on the table, so it's all depending on different circumstances. But there's like several different aspects of how it's gonna recognize it."* Overall, AI-enabled privacy techniques create a knowledge imbalance where blind users cannot directly verify its outputs [6, 97, 118, 119], and assessment descriptors may still maintain these asymmetries. This reflects a broader issue of enforced trust, where blind people are often asked to trust and accept outcomes without questions. As P23 argued, *"there's that whole trust business. And we're asked to do that a lot as blind folks. Trust the sighted folks, trust the AI, trust the public transit, trust me, that and the other thing."*

*4.1.2 Unclear Purpose of Assessment Descriptors.* Some participants perceived assessment descriptors as redundant. Despite framing our questions and interest in designing descriptors to verify obfuscation, some participants considered certain assessment descriptors an extension of their original visual task, not a way to evaluate the quality of privacy features. For instance, participants thought of measurement descriptors as either irrelevant to the task, or as re-stating known information. Specifically, in the first audio probe, some participants questioned how receiving dimensions of the microwavable meal would enable the hypothetical user to proceed with their cooking task. As P12 said, *"I already know the size of what it is I'm holding."* P3 particularly thought that providing such information could be infantilizing. They explained, *"I'm holding it in my hand like that's kind of belittling like 'Oh you don't know what the size is?'"*

In the group discussions, even when a few participants explained how assessment descriptors are different, many participants did not find describing measurements and color useful. We observed this play out in both FG4 and FG5. P2 explained that measurements could be *"almost like a safety net [...] I have those dimensions as a fallback in case it doesn't adequately assess what that object is."* However, P8 responded to P2 with a counter that objects can have the exact measurements. They said, *"that box could be like a box of waffles and not a microwavable dinner [...] the dimensions do not identify the product."* Different objects may have similar sizes, rendering measurement-centric assessment descriptors insufficient. Overall, providing features without context for how they should be interpreted or specificity in how they relate to the visual content made it difficult for participants to use them for reasoning about AI behavior.

*4.1.3 Navigating the Complexity of Human-Enabled VAT.* Participants pointed out that designing privacy features in human-enabled VAT is especially important as they perceived higher privacy risks than AI-enabled applications [4]. However, participants asserted that obfuscation techniques and assessment descriptors could be complicated given the dynamic nature of video calls compared to static images. P20 explained, *"each time I move the camera, [the obfuscation] might need to be readjusted, and what I want to be hidden will show up for a little while and then disappear again."* Alternatively, P20 suggested pausing the video stream periodically and checking in with users. They explained, *"[privacy features] could tell me that it is ready to restart the recording and ask 'Do you want to restart the video' or something similar as a warning"* (P20). Furthermore, balancing assessment descriptors and guidance from sighted people could be cognitively demanding. As P16 explained, *"if my phone was trying to help me focus on something and it is chatting when there's a sighted person on the other end who is also trying to help me. I think these two things would clash."* It may be further challenging as blind users may need to adjust their cameras and change certain light conditions to increase visibility for sighted people [6, 17]. Instead of in-situ privacy tools and assessment descriptors, P7‡ proposed that privacy features should be applied before using human-enabled VAT. They said, *"I think kind of it would have to be a layer that you go through before you connected to the volunteer."* In essence, assessment descriptors could be distracting and inaccessible during the dynamic nature of video-based interactions, particularly human-enabled VAT. Participants offered some suggestions, such as pausing videos and performing privacy modifications prior to engaging with sighted people in human-enabled VAT.

*4.1.4 Assuming Prerequisite Knowledge.* Assessment descriptors assume users already know what the visual properties of an object should signify. Yet, that is not always the case, especially for users born blind who may not find visual aspects particularly meaningful. For example, when sharing how far away a certain object is from the user, P23 challenged the utility of distance as an assessment descriptor for people who are born blind—noting, *"I didn't get a ruler and measure that [...] for someone who has previously had vision and has a bit more of that spatial awareness that might work. But for those of us who have never seen, that's absolutely not going to do it"* (P23). Similarly, when an object's true color is unknown, it becomes less useful as a way to confirm AI functionality. P20 said *"I might not know the color of the box."* While blind people may know object colors in some cases (i.e., the color of familiar objects or colors from previous sighted experience), perceptions can vary based on when a person became blind. P6 explained, *"I lost my sight later in life, so I understand color. But to others it may not be beneficial."*

Short and vague assessment descriptors, such as, "approximately 8 inches in length and 4 inches in width," do not contain enough information for blind users to verify obfuscation. Participants recognized that assessment descriptors, such as distance, are relative. P3 explained *"it would drive me crazy [...] what is this 8 by 4 referring to?"* Participants wondered whether the size information referred to the object dimensions or the particular region of interest (i.e., cooking instructions). In our study and past work [6], visual privacy tools can be used as a filter mechanism to spotlight specific aspects. For example, P4 shared how obfuscation could be used to solely *"to focus on the veggies"* section of a menu. Spatial assessment descriptors become confusing when the dimensions referenced are not linked to a particular region. Furthermore, participants commented on how distance and dimensions may change based on the camera's position. P20 explained, *"I assume it depends on the distance between the camera and objects, the object might look smaller or bigger."* Communicating distance without any information on reference points is inadequate.

Even with the additional details, some participants discussed how color, distance, and dimensions as descriptors could be *"sighted-centric"* (P23), misaligned with how blind people navigate visual contexts and attempts to propagate sighted ways of thinking through centering visual elements that are not relevant to blind people. P23 explained how this is endemic in the design of assistive technologies:

> *"I think that it comes down to something that the sighted world hasn't figured out that it does yet, and that is that of all the disabilities, blindness seems to be the one that the world fears the most. And so many people cannot even begin to figure out how they would do things on their own without being utterly dependent on technology or another person as a blind person. So they attempt to design these things with the best of intentions, but they don't have a concept of how blind people live day to day."*

Participants are critical of assessment descriptors, noting how they could miss important information, conflict with VAT, produce inaccurate output, and reflect sighted norms. Despite their limitations, some noted that assessment descriptors are *"better than nothing"* (P20) and, if redesigned, they have the potential to be beneficial.

## 4.2 Preferred Assessment Descriptors

In this section, participants shared their perspectives on assessment descriptors that matched their everyday experience with VAT. Moving away from providing assessment descriptors such as location and color, participants wanted to know the object's name and unique features. Beyond the object of interest, participants argued that receiving the description of surrounding objects may help detect errors by cross-referencing with their familiarity of a particular space. Lastly, participants speculated on the benefits and limitations of incorporating haptics and audio tones instead of verbal assessment descriptors.

*4.2.1 Just Name the Object, Its Distinct Aspects, and Provide the Option of OCR.* Participants shared that emerging privacy tools should identify the object of interest and its unique visual characteristics. Object descriptions could identify the object (e.g., "microwavable meal" or "delivery box"). Furthermore, descriptions could also include unique properties like *"the shape [...] the material it's made of"* (P20). For instance, assessment descriptors could convey the presence of a particular logo like *"Amazon box"* (P23) or that even if it is *"something that has text on it"* (P24). Rather than designing privacy features that prompt blind users to identify specific objects of interest and then generate assessment descriptors to validate detection, participants suggested that privacy tools should name the objects and offer options to highlight unique attributes.

However, this approach would not necessarily solve the inaccuracy issues discussed in 4.1.1, especially given how object recognition techniques performs poorly on images taken by blind people [53, 77] and for non-Western objects [7, 37]. As P24 pointed out, *"of course, you have to hope that you're not getting any erroneous information."* For that reason, some participants thought OCR could be helpful in objects that contain text. P9 said, *"if I start hearing something about microwave [...] that would give me confirmation that what it's looking at is what it's gonna select."* P22 agreed, noting that they would use OCR to look for keywords. They elaborated, *"like prepare oven door, things of that nature [...] Then, I'll know I am in cooking directions. But if you tell me like saffron, I get a feeling I'm in the wrong area."* Nevertheless, OCR also has it limits and *"blurts out garbage and you're just trying to piece it together"* (P22) when the packaging has complex designs such as *"it's written in something fancy or like if it's inside a graphic"* (P8). One participant raised concerns about false positives during a high-risk and privacy-sensitive case of trying to obfuscate social security numbers. P19 shared, *"there could be numbers that have the same format as a social security number. Maybe a date of birth that is misprinted, or maybe the OCR is trying to pick up on some kind of certificate number."* Accordingly, earlier during the interview, P19 suggested that OCR should accompany image quality indicators to verify output and retake the photo if it is blurry. Specifically, P19‡ said, *"I can just picture the little robot voice: 'Here's the text, but I'm not sure. You might want to double check. You might want to take the picture again.'"* Overall, participants anticipated the benefits of OCR to aid in accessible error detection yet noted the need to incorporate quality notices.

*4.2.2 Describe the Entire Scope.* Instead of providing assessment descriptors for one object of interest, some participants suggested describing multiple objects within the environment to assess quality. In essence, participants speculated that if assessment descriptors would include more than the specific object, they could validate accuracy based on their familiarity with a space. For instance, P20 explained, *"let's say we are looking at a bunch of different things in the kitchen [...]. So the [focus mode] can tell me: 'The camera sees, you know, this, this and that, which one you wanna focus on?'"* If assessment descriptors included numerous objects in P20's kitchen, a space they already know, it would give P20 *"more assurance"* as they could examine what privacy tools were able to recognize and what it might fail to capture.

Additionally, providing an overview of the space could prevent accidental disclosure of private content, a key concern identified by blind users in prior research [99]. P10 explained, *"sometimes we forget that we put things there just because. You know, when you don't see it, you forget [...] Out of sight, out of mind."* For instance, P4 said *"I took this nice picture, and I sent it to my friend. Later, I realized that it also captured the sanitary pads which I didn't want it to include in the picture."* Thus, expanding assessment descriptors beyond one specific object of interest could help participants find potential privacy leaks, and apply obfuscation or remove private elements from the space. Some participants noted the opportunity to include *"an warning kind of thing like 'heads up like all this sensitive information.'"* (P7). However, a few participants complicated this approach of nudging for sensitive content. Corroborating past research [6], participants echoed that privacy is *"really subjective"* (P25‡), and some questioned the benefits of receiving constant privacy notices. Counterintuitively, P23 explained that privacy tools could decrease user control. They said, *"what it's doing is taking away the power of choice. It is deciding for you what the item is based on your keyword criteria. And as blind folks, we get our autonomy taken away rather often. And that's a turn off."* Similarly, P3 added:

> *"There's an element, whether we intended it to be so or not, of censorship [...] if you start like cueing people, for example, there's potentially some personal information, you're almost inducing or like pushing the user to hide that or to react in some way. [...] Maybe you have medication and you need to know the dosage and how much to take because this is the first time and you didn't get to talk to the pharmacist. So you need that information, right?"*

In sum, participants discussed the value of including several objects in the users' environment, not just the object of interest. By referencing the detected objects in the space with their knowledge of their surroundings, some participants anticipated that it might prevent unknown privacy leaks. Yet, there are drawbacks to this approach as there is not a universal definition of sensitive content, and it could overly influence users to make certain decisions that they otherwise would not have preferred.

*4.2.3 What Could Descriptors Sound & Feel Like?* The assessment descriptors we explored were presented verbally, and some participants proposed other sensory experiences. Reflecting on interaction techniques in VAT, some participants enjoyed the haptic features of some VAT applications. For instance, P9 valued that OneStep Reader[4] uses haptics to guide blind users in taking accurate images. They said, *"when [the camera is] maybe tilted more to the right, I think it like vibrates more to the right, and then If I tilt it more centered, I could feel a vibration depending on which way it's aligned."* However, understanding haptic feedback requires a learning curve. P6 explained, *"not everyone understands haptic; I still barely understand haptic."* Because there is an endless array of haptic cues that users need to interpret, P24 advocated for including blind people throughout the design process. They explained, *"a blind person who understands a little bit about the idea of haptics needs to at least be on the research and development team [...] we need to be involved in the actual research and development, not in the testing afterwards"* (P24). Some participants brought up the hardware limitations of haptic feedback. P20 elaborated, *"maybe if I have a huge Braille display type of thing [...] But I mean those are rare and expensive. So not a lot of people will have those."* In another focus group, P23 argued that haptic feedback may be more fitting for a wearable (e.g., Apple Watch) rather than a mobile phone. They asserted, *"getting like haptic feedback when it determines where your hand is located and it recognizes the item. That would be delightful on an Apple Watch or on a comparable device. But on a phone that's not going to do it."* Haptic feedback may improve assessment descriptors but could be cognitively demanding and technically challenging.

Some participants discussed the opportunity to incorporate audio tones. Similar to how some screen readers (e.g., Jaws) include audio schemes to distinguish between various web elements [93], P10 wondered if assessment descriptors could include *"like they would like on Jaws: when it's a capitalized matter, or something has a higher pitch."* P17‡ provided an example of a potential use case and how audio tones could be beneficial. They elaborated, *"say you're in the bathroom, you drop the bra on the floor, you don't know where it's at, you're trying to feel for it, but don't want [VAT] to know you're in the bathroom looking for a bra on the floor. It might just zero in general areas where it will block out the toilet, block out the sink, block out the tub area, it will just show the floor and then it will guide you to where it's at. If it's to the left of you, it might be a high pit beep, beep, beep, beep to the left to go to the left. [...] Something creative to that point"* (P17‡). However, like haptic feedback, audio tones require a learning curve. P20 explained, *"there are these audiographing apps. So you hear the graph. But it takes a while to get used to it, like. Oh, the X-Y access is starting from here. After a while you get it, but still. Not perfect. And it's a long process to get used to it."*

[4]OneStep Reader is an AI-enabled VAT [11, 54].

## 4.3 The Missing Pieces: What Assessment Descriptors Still Need

The previous sections analyzed the types of assessment descriptors that (mis)align with blind users' process of detecting errors. In this section, we detail broader considerations to support the design of assessment descriptors and privacy technologies. Specifically, participants emphasized the need to offer customization, provide in-depth explanations on how AI-enabled privacy tools are created, and facilitate support and training sessions.

*4.3.1 Customizing Descriptors Type, and Depth.* While participants suggested some assessment descriptors in section 4.2, a universal approach remains elusive since assessment descriptors are *"contextual"* (P22) and *"not all blind people are the same"* (P23). Reflecting on their experience using AI-enabled VAT, participants shared that *"these apps already do a pretty good job of identifying things, but getting the additional information is where they fall short"* (P8). P4 elaborated that Seeing AI, an AI-enabled VAT, *"just added that feature [...] there is a more information button"* to receive additional insights. Similarly, some participants wanted to gain more in-depth assessment descriptors. As P2 noted:

> *" I'm having a mental block of trusting it enough to say that it can tell me what the item is. And that's why I think I would prefer to know that extraneous information probably after if I needed it [...] if you had a button that wouldn't tell you it initially, but you could press and it's almost like more information about the product."*

Having the choice to receive more detailed assessment descriptors enables participants to *"find the right balance of cognitive load"* (P7), especially given the dynamic nature of using visual assistance technologies (VAT) [6]. Additionally, allowing users to customize the depth and type of assessment descriptors *before* use could be beneficial. P6 argued, *"when you're actually activating the focus mode [privacy features]. It would have those options before you start it. Like this is how much information I want, this is what I'm looking for, all that in the menu, and then you hit did start or whatever."* A few participants imagined playing a more engaged role in shaping assessment descriptors based on their needs. For example, P9 said *"[I] could like train [privacy features] so in my office I only want [it] to see these items."* Accordingly, participants envisioned a fluid approach to descriptors where they would change according to the context and content. In essence, *"visual interpretation is so subjective"* (P24), participants stressed the diversity in blind communities and advocated for options to customize the level of details.

*4.3.2 Incorporating Data Transparency.* Assessment descriptors are a small part of detecting errors in AI-enabled privacy tools. Before evaluating a specific output of obfuscation, participants wanted to understand the structural components of these techniques as a way to cultivate trust or engage in critical refusal [48]. In particular, they wanted to understand *how* privacy features process and collect data. P2‡ shared with us a hypothetical scenario of wanting to obfuscate a condom packet, noting that they would like answers to the following questions: *"how does it know to code this information? How does it read the brand without necessarily getting sidetracked by the plastic thing of that product?"* Participants also wanted to *where*

data processing occurs, whether locally (on-device) or externally. P8 explained, *"it's getting processed locally on my phone or computer?"* Similarly, P3 added *"how much of the information is being saved or used for other reasons because these things don't just happen out of a vacuum, right?"* Furthermore, participants wanted to know *what* types of data are used to train AI-enabled privacy tools. P22 elaborated that obfuscation developers should answer questions, such as *"how much do you trust your data at what level and where your sources are from?"* Similarly, P7‡ said, *"if I knew that something was three-quarters powered by Twitter and one-quarter powered by Wikipedia I would probably trust it a lot less than something that was maybe three-quarters Wikipedia and one-quarter Twitter."* Instead of focusing on assessing particular outputs (what tools like assessment descriptors and confidence scores aim to support), participants desired a more holistic view of obfuscation and posed several questions that would inform their (non-)use of obfuscation.

During focus groups and interviews, participants further debated what meaningful transparency in obfuscation could entail. They expressed differing opinions on the value of detailing the edge cases of AI-enabled privacy tools. Some participants wanted to understand the instances where privacy technologies may fail given that *"under all these other adverse conditions, including bad weather, the performance might be nil or anything between"* (P22‡). P8 said, *"what are the limitations of the focus and background options?"* They wanted concrete and specific types of failure cases. For example, P18‡ posed the following question: *"when we have people like the stupid guy from Twitter [Elon Musk] who named his child with a symbol. Does our AI know that it's a name?"* In essence, some participants were interested in examples of errors in AI-enabled privacy tools. However, others did not trust that developers would provide detailed information on the limitations of emerging privacy features. As P22 ‡ explained, *"you're not gonna get that honesty out of a company. [In my experience,] there is no way that I would put in writing any limitations of my product that weren't mandated."* Drawing from their experience with disclaimers in AI-enabled VAT, such as *"Seeing AI when they come up with a new thing, it tells you a little bit about like 'please use with caution this is an experimental feature'"* (P19), some participants did not find generic information about inaccurate use cases helpful. For instance, while Seeing AI indicates that certain features are under development by noting that they are a "preview" [62], some participants did not know what this meant and how preview or experimental features related to accuracy. When asked what they thought "preview" indicated, P5‡ said, *"I never figured out what to do with that. I just flick up or down until I get back to it, and then it doesn't say preview anymore."* In contrast, others noted that it meant *"basically in beta"* (P9) but were unsure about specific limitations. Ultimately, current disclaimers and notices about inaccuracy may be perceived as *"more of a cover your butt kind of thing"* (P3). Participants also noted that users do not read these disclaimers, and they are often placed in Terms & Conditions. P3 added, *"who actually reads that? [...] it's almost like one of those situations where it doesn't matter if you do or not."*

Even outside of errors and disclaimers, some participants pointed out general transparency issues with VAT. Notably, participants mentioned the incident where Be My AI would not process visual content with faces. P1‡ shared with us, *"It was actually rejecting any images that had people at all even if it was a magazine or a poster on the wall."* Users were not notified about these guard rails until much later in a public blog [13]. For paid VAT apps, such as Aira, participants noted that they increased prices without consulting the community. P5‡ explained, *"essentially the blind community blew up the Aira phone line because they said, 'for you guys to get the new pricing you have to call Customer Care.' So, everyone would call customer care and no one could get through."* These practices, in addition to several privacy violations participants raised in our study and prior work [6, 97, 98], led some participants to distrust VAT. As P14 asserted, *"we are so vulnerable; we are taken advantage because our disability."* In relation to developing privacy technologies, one participant speculated about the benefits of designing AI-enabled privacy tools outside of VAT, rather than a feature within VAT. Specifically, P7 noted, *"this would be working as a go-between or a kind of filter layer. Before you engage in something like Seeing AI or whatever. It is used to narrow focus on what your camera had in its scope prior to launching an app that would be sending stuff to the web."* Having AI-enabled privacy techniques separated from VAT applications enables blind users to overcome the challenges of trusting VAT to be accountable and transparent. P7 elaborated, *"I'm not really interested in the politics of who's AI security I want to trust more as much as I am interested in the conversation of how can the data I don't want to be sharing with the cloud [be safeguarded]."*

*4.3.3 Providing Training Materials & Human Support.* Participants anticipated the benefits of receiving and developing training materials on how to use privacy features. In line with prior work on the challenges of non-visual camera aiming [7, 64, 71], some participants suggested incorporating camera guidance when building accessible obfuscation. For instance, P3 noted the value of *"instructions on how to put something in focus mode [...] Like, in Seeing AI, where you have a document, it'll say hold steady when it's got it, but if it says like left edge not visible. I'm being cued to then move to the left more."* In addition to advice on camera aiming, participants wanted to know tips and tricks on optimizing these emerging privacy features. P8 said, *"if we're going to start relying more on AI, then we also need to learn how to interact with it."* P20 imagined creating and maintaining a community online forum where blind users would document their experiences using privacy features. They elaborated, *"[if a blind user] wants to use [privacy tools] for this, they can go there [online forum] and check out if it works for that scenario from other people's experiences before they even try, because if it something very serious and important they may want to know beforehand if it works or not. "*

Training is especially important for blind people who are born blind and may not have an understanding of visual concepts like focal points or backgrounds. P7 explained, *"blind consumers is a very wide range of pre-existing knowledge bases, whether it comes to technology, whether it comes to visual concepts, so ensuring that there are sets of information for users who have a very limited conceptual or technical background to understand the basics."* P16 added that there is an opportunity to have blind individuals create these training materials for their community *"because they can explain it in a way that other blind folks can understand."* Indeed, blind people already have community support to help each other with technology. For instance, P10‡ said, *"in the blind communities, we talk about all the*

*stuff we use. We share what app is better to use than some, and then some people give you certain tips."* Some of our participants indicated that they create technical resources for the community. P18‡ shared with us, *"I do teach people how to use Seeing AI."* P3 added, *"I've worked with folks on how to use these apps, and we've always taught them how many you want it so many inches from the camera. Because if it's right, if the counter is right on the object or right on the text you're not going to get all of the text."* In sum, participants advocated taking a community-centered approach when developing training materials for using privacy tools.

However, participants recognized that training materials were not enough, and some wanted the option to receive support from VAT in cases where they were willing to share their data. P7 explained that data could be optionally stored *"on device and not on the cloud, but on device or somehow otherwise encrypted. Every session where both the original photo and the resulting photo are available to the user to use for reporting."* They imagined that a *"live chat feature would be the best feature"* (P10). Participants were particularly against chatbots and preferred human support. P6 explained, *"the human element is being taken out so quickly [...] [chatbots] all depends on the wording and they're going through an algorithm of typical responses from frequently asked questions. But you may have a question that does not quantify within the parameters of what is already installed within the chatbot's algorithm."* Chatbots, especially role-based chatbots, are only able to respond to a limited set of questions and requests. Participants predicted that chatbots would likely fail to support privacy features. Even AI-based chatbots have their faults, which could include harmful bias. P23 explained, *"I actually had a conversation with ChatGPT [...] AI does some really nasty stereotyping of blind individuals and what they can and cannot do. I don't want my technology support to be solely based on AI."* Ultimately, participants desired a *"a human who can stop and think about what I'm saying and really show that they are understanding by actively listening, by taking down the information and then by clearly communicating"* (P9). Taken together, these findings indicate enthusiasm for co-creating tutorials on using privacy tools and providing support for errors.

# 5 Discussion

Our findings revealed the possibilities and limitations of using assessment descriptors to support non-visual error detection in obfuscation. We found that some assessment descriptors, such as distance, color, and dimension, are insufficient. Alternatively, participants discussed assessment descriptors that name the object and its unique aspects, describe surrounding objects, and incorporate multimodal techniques. Beyond specific assessment descriptors, participants emphasized the need for customizable tools, greater transparency around how AI-enabled privacy systems are developed, and access to training and support when errors occur.

In this section, we build from our findings to inform two domains: (1) accessible verification of AI tools and (2) emerging privacy techniques. Particularly, we examine how sighted bias shaped the framing of certain assessment descriptors. Then, we situate participants' preferred assessment descriptors within the existing literature on visual description. We also illustrate opportunities to improve obfuscation by developing support avenues and balancing the tensions of privacy alerts. Finally, we offer directions to improve assessment descriptors, obfuscation, and VAT moving forward.

## 5.1 Accessible AI Verification

Our analysis offers insights and design directions for accessible AI verification (i.e., assessing AI output). While AI technologies are rapidly deployed for visual accessibility, supporting blind people in finding instances of errors is overlooked in existing literature [1, 7, 50, 74]. Particularly, prior work identified that blind people perceived that AI-enabled privacy techniques are not only prone to error, but also lack accessible means of detecting errors [6, 97, 118]. To examine and design for verification, we explored audio probes of assessment descriptors with blind participants. Drawing from our findings, we highlight three key takeaways to inform the future of accessible AI verification. Particularly, we reflect on how sighted bias influences the design of some assessment descriptors, and compare assessment descriptors with existing visual description standards.

*5.1.1 Challenging Sighted-Centric Assessment Descriptors.* Assessment descriptors that we initially introduced are, as P23 eloquently described, *"sighted-centric"* (refer to 4.1.4). Reflecting on our position as sighted researchers and this study's findings, we wondered: What makes describing visual properties sighted-centric, and how can we push back against sighted norms? In general, sighted-centered design is a set of technologies and practices that privilege sighted sensemaking and marginalize blind and non-visual ways to relate to the world [92, 103]. To further understand how sighted sensibilities are centered in our articulation of assessment descriptors, we situate our findings in disability studies and accessibility scholarship. In particular, we identified two cases of sighted bias: (1) using visual-centric terminology and (2) limiting assessment descriptors to one modality.

First, **spatial (i.e., related to sizes or distance) and color-based assessment descriptors rely on sighted language.** In general, creating technologies to enable blind people to navigate indoor and outdoor settings has long been a topic of interest in accessibility and AI research (e.g., [40, 83, 87]). Yet, the sighted norms that permeate describing directions are rarely questioned [106, 112]. In particular, Siegfried Saerberg, disability studies scholar, argues that providing directions is not objective and is shaped by various subjectivities [92]. Saerberg [2010] writes, "[sighted people] forget that 'straight ahead' is not self-explanatory. Very few sighted persons are able to move beyond such assumptions" [92]. For instance, using a ruler to measure a book and stating its dimensions might seem like a purely factual task. However, disability studies scholars and philosophers argue that space is not fixed but continuously negotiated and reshaped. Most related to work, Vincenzi et al. [2021] describe the rupture and repairs of sighted and blind pairs, such as when sighted guides suddenly leave or fail to communicate directions [106]. In our study, participants detailed how spatial assessment descriptors lack enough details to validate privacy tools, noting how distance and dimensions are vague and relative (i.e., based on how far the user positions their camera). To reframe sighted language, participants in our study provided suggestions for more accessible ways of communicating spatial assessment descriptors, including specifying what the dimensions

refer to (i.e., the object of interest or scope of obfuscation) and incorporating reference points of interest. Additionally, color-based assessment descriptors reproduce a sighted worldview that homogenizes blindness. As noted in 4.1.4, participants who were born blind did not find some assessment descriptors, such as color, memorable, reducing any potential utility to cross-check obfuscation.

Second, **our conceptualization of assessment descriptors began with one modality: verbal descriptions. However, participants discussed the benefits of incorporating haptics and audio tones.** While on the surface, designing assessment descriptors to be verbally described may not appear sighted-centric. Nevertheless, it does reflect a lack of lived experience with visual access. Rod Michalko, disability studies scholar, describes the start of acquiring a visual disability as "inescapable epistemic contingency," denoting ongoing negotiation [81]. Blind epistemology (ways of knowing) is fluid and relational, shaped by objects, environments, memories, and other people [17, 90, 111]. In contrast, sighted epistemology is dependent on vision [92], a sensory stressed from a young age (e.g., "watch out" and "look both ways") [103]. In 4.1.3, participants shared how using VAT, especially when it involves remote-sighted volunteers or crowd workers, can be mentally demanding. They also highlighted how assessment descriptors can contribute to cognitive overload. Similarly, blind participants in Alharbi et al. [2022]'s study noted that obfuscation (without assessment descriptors) is *already* cognitively taxing [6]. Subverting a sighted-centric emphasis on singular modalities, designing for accessible AI verification may entail moving beyond verbal cues. Our findings suggest the potential for taking a multimodal approach, incorporating audio tones and haptic feedback to negotiate obfuscation errors. However, our participants acknowledged the complexities of this design space, from infinite audio tones and hardware limitations to a steep learning curve. There is an interesting opportunity for future work to draw from our study and past work [6, 65] to explore the possibilities and pitfalls of using multimodal approaches to verifying AI outputs, particularly in privacy-preserving techniques such as obfuscation.

Overall, framing assessment descriptors and, more broadly, describing visual elements is a value-laden activity. If done incorrectly, it could be reflective of sighted preferences. Bennett et al. [2020] introduced the concept of "non-innocent authorizing of care" to explain how sighted people hold the authority to describe visual elements, choosing what to emphasize and what to leave out [20]. Designing visual descriptions, including in high-risk cases like validating privacy techniques, is entrenched in power dynamics. Reflecting on our study, the assessment descriptors we introduced in our audio probes center around sighted sensibilities, privileging distance, color, and dimensions, disconnected from how blind people identify objects and find AI errors. While we aimed to introduce these assessment descriptors to inspire discussion, including the refusal of such assessment descriptors, we acknowledge the harm in reproducing sighted norms, and we are grateful that participants actively challenged us.

Taking a broader view, we argue that the field of accessibility must collectively examine and work to dismantle sighted bias and centrism. Within visual accessibility scholarship, there is a growing interest in the subjectivity of visual description [18, 31, 106], including investigations into why crowd workers produce conflicting descriptions [23] and creating visual description standards with blind users [96]. Beyond these individual-level analyses, researchers also focused on structural dynamics, particularly the power differentials between researchers and community members [20, 113, 114]. For example, Williams et al. [2023] critiqued the disengaged and interventionist tendencies of HCI and assistive technology research, advocating instead for a "counterventions" approach that emphasizes self-critique and participant agency [113]. Reflecting on our work, we recognize that early formative studies may have helped identify sighted-centric framings and allowed for course correction. Still, we hope our analysis offers valuable lessons for future research. Moving forward, there is great promise in developing methods and toolkits that enable researchers to interrogate their research questions and methodologies reflexively. We are encouraged by, and eager to contribute to, the emerging conversations in this space [78, 100].

*5.1.2 Assessment Descriptors and (Dis)Connections with Image Description.* While participants were critical of the assessment descriptors shared in focus groups, they did suggest more accessible assessment descriptors. Participants' assessment preferences can be traced within the broader discourse on alternative text (alt text). In some ways, our participants' preferences for framing assessment descriptors align with existing standards for writing alt text [85, 91, 96, 108]. For instance, in 4.2.1, participants valued object recognition approaches that directly named the object of interest (e.g., microwavable meal) and coupled with OCR if the text is available. Similarly, Stangl et al. [2020] studied blind people's preferences for image visual description [96]. In describing images containing objects, they found that blind people wanted to know texts, names, materials, colors, and logos/symbols. Likewise, our study participants wanted assessment descriptors to include distinct features of an object of interest. Nevertheless, participants' views of meaningful assessment descriptors sometimes diverge from existing alt text guidelines. As previously noted, color was the least desirable assessment descriptor in our focus groups, whereas including information about colors could be particularly useful when visually describing fashionable outfits [31] or makeup palettes [72]. Spatial assessment descriptors were also heavily critiqued. However, when writing alt text for data visualizations, describing the dimensions of different regions in a graph and how they compare could be useful to blind users [107]. In essence, assessment descriptors are a type of visual description. While we understand how to generally write alternative text in different domains, describing content as a way to enable blind people to refute or accept AI is largely understudied.

Assessment descriptors are a type of visual description, but they differ from the original task blind people intended to use VAT for (e.g., object identification). The designers of obfuscation tools should differentiate between these various interactions. For instance, if a blind user wants to use VAT to read mail and enable obfuscation with assessment descriptors, the original VAT task is reading mail. Considering the close resemblance and constant negotiation between these visual tasks, participants' first reaction was to perceive assessment descriptors as part of the original visual task. For example, when we played the audio probe for focus mode, some participants were confused about how or why dimensions would help the hypothetical user read cooking instructions (refer to 4.1.2). Accordingly, participants found some assessment

descriptors repetitive and disrespectful since they already knew this information. Future work should develop and study user experience (UX) writing that distinguishes between visual information for the original visual tasks and assessment descriptors. One potential direction could be explicitly framing assessment as a question and highlighting the privacy-enhancing action. For example, developers could explore the following language: "this appears to be an object with [assessment descriptor(s)]. Is this what you would like to [spotlight/obscure]?"

## 5.2 Emerging Privacy Tools in VAT:

Building on our findings, we emphasize the importance of enabling blind people to use obfuscation, and balancing the harms and benefits when notifying blind users of potential private content in obfuscation techniques.

*5.2.1 Supporting Blind People to Obfuscate Content.* In 4.3.3, participants voiced the need to co-create instructions on using AI-enabled privacy tools like obfuscation with blind users. Our finding differs from a prior study that found that most of their blind participants could use an obfuscation prototype easily without formal training [118]. The difference between our study and past research could be attributed to the nature of the visual content examined. Previous research that reported ease of use investigated how blind users would apply obfuscation on mostly researcher-provided static images [118]. In contrast, our study focuses on VAT, which involves blind users taking static images or videos. Building and extending our analysis, future work could construct teaching scenarios that involve different objects (e.g., objects, paper documents, or photos) and environments (e.g., home or workplace) in collaboration with blind communities. Notably, such learning content should respond to diverse technical experiences and visual disabilities. For instance, tutorials may include a hands-on introduction to visual concepts like obfuscation, blurring/blocking, background, and foreground.

In addition to training materials, participants discussed the value of receiving human support during obfuscation errors. Extending prior work that advocated for safety mechanisms [6], participants wanted the option to share data and receive human support. They strongly reject automated approaches to support, such as rule-based or AI-powered chatbots, citing concerns of bias and lack of nuanced understanding of accessibility. Instead, participants advocated incorporating human support to relay feedback or concerns to development teams. Nevertheless, including human-based support to detect and resolve obfuscation could introduce additional privacy violations. Accordingly, support employees must be subject to confidentiality agreements, and support sessions should not be recorded or securely stored.

*5.2.2 Understanding the Tensions of Privacy Alerts.* In our study and prior work [6, 97], blind people identified that a key limitation of obfuscation techniques is the requirement to know of private content first. Blind people need to be aware of the presence of private content within an image or video, to then use obfuscation techniques. This could be difficult or unrealistic since blind people may accidentally or unknowingly capture private content [99]. Accordingly, some participants in our study highlighted the importance of alerting blind users to potential private disclosures as a way to prompt the employment of obfuscation tools. However, a few participants pointed out the potential harms of nudging blind people to use obfuscation and redact suggested private content, raising concerns over the pressure of obfuscating objects that are not regarded as private to the user. Particularly, P3 described these types of nudges as coercive (refer to 4.2.2). Similarly, in prior work by Alharbi et al. [2022], many of their participants objected to automatic obfuscation decisions and characterized alerts to obfuscate content as intrusive [6]. Broadly, HCI scholarship has critiqued nudges for being potentially manipulative and lacking transparency [33, 109]. Our finding resurfaces and underscores the need to balance the utility of privacy alerts with blind people's sense of agency and control over obfuscation techniques. Particularly, future research could further explore what it might mean to notify blind people of potential private content while preserving blind people's right to obfuscate or not. Specifically, an upcoming study may investigate if and how reflective privacy alerts that induce "friction" [35, 80], allowing users to pause and think rather than immediately prescribing obfuscation, could balance the tensions and opportunities of obfuscation nudges.

## 5.3 Design & Research Directions:

This section offers takeaways and directions for future research and VAT. Particularly, upcoming work may explore the possibilities of co-creating a typology of assessment descriptors, incorporating Visual Language Models into assessment descriptors, building community-centered tutorials on how to use obfuscation, and supporting transparency in obfuscation and VAT.

*5.3.1 Co-Developing a Typology of Assessment Descriptors.* As discussed in 5.1.1, the assessment descriptors we explored in this paper are influenced by our sighted bias. When developing techniques for validating AI results, upcoming work could audit for sighted-centric language and engage with blind communities, particularly varying the range of visual memories, to develop more accessible assessment descriptors. Specifically, by drawing parallels from existing literature and our findings, future work could explore creating a typology of desired assessment descriptors with blind communities. For instance, using datasets of content perceived as private by blind people [4, 6, 52, 99], upcoming research may co-produce assessment descriptors for different objects with blind communities.

*5.3.2 Navigating the Dual Edge of Visual Language Models.* Participants argued that the presented assessment descriptors are rigid and vague, limiting their potential to support error detection. With the recent advancements in Visual Language Models (VLMs) [73, 117], future work could enhance assessment descriptors with VLMs and examine the effects of contextually rich and customizable assessment descriptors. However, unlike on-device lightweight models, VLMs require off-device processing, which entails privacy and security risks [36, 70]. Furthermore, prior work has described Large Language Models, a component of VLMs [73], as "bullshit machines" [57] or "stochastic parrots" [16] to highlight how they are designed to emulate confidence rather than provide factual information. Accordingly, this limits the promise of VLMs in assessment descriptors.

Upcoming work seeking to investigate VLMs in assessment descriptors, or AI verification more broadly, should consider and mitigate accuracy and privacy dimensions.

*5.3.3 Establishing Systematic Transparency in Obfuscation as a Precursor to Verification.* Prior work, including our own, has focused on adding additional information, such as the visual descriptions or confidence ratings, to support blind users in validating AI [5, 59, 60, 74]. That said, our results highlight the importance of adopting a broader perspective, informing users of the inner workings of AI technologies. Departing from assessment descriptors, participants offered suggestions to increase transparency of AI-enabled privacy features and VAT. Particularly in 4.3.2, participants expressed a desire to (1) understand how obfuscation systems are designed, (2) learn about how AI systems are trained, and (3) identify scenarios where obfuscation might not catch or conceal private information. The vast majority of prior work on AI transparency has focused on engineers and data scientists (e.g., [41, 49, 66], with comparatively little attention given to user experiences [22, 104] and even less to accessibility contexts [1, 7, 74]. Our findings lay the groundwork for further exploration in this area. Building on these insights, subsequent work could focus on co-developing obfuscation transparency guides in collaboration with blind communities, exploring strategies for clearly communicating AI limitations, and using language that balances technical accuracy with varied levels of digital and informational literacy. Nevertheless, some participants were hesitant to trust obfuscation transparency guides. Certain transparency efforts can sometimes serve as a form of "ethics-washing," creating the illusion of high ethical standards while the actual practices do not reflect such principles [24, 115]. In the next direction, we will explain why participants are critical of the transparency guides and articulate steps toward meaningful transparency.

*5.3.4 Improving Transparency in VAT & Building Obfuscation Tools Outside of VAT.* Some participants reasoned that they are reluctant to trust transparency guides for obfuscation because a couple of key VAT have opaque practices when communicating AI output and changing pricing policies. For instance, as shared in 4.3.2, Seeing AI described experimental features with the confusing label of "Preview." Practically, VAT could indicate features under development using a more explicit title, such as "undergoing testing and improvement" or "experimental." VAT may go further by articulating why this feature is still under construction and providing cases where it could produce inaccurate results. Further, because participants' apprehension of obfuscation stemmed from a lack of transparency in VAT, P7 (in 4.3.2) suggested creating privacy tools outside of VAT instead of incorporating them as a feature within VAT. Having decentralized privacy techniques that are disconnected from a particular visual assistance technology enables blind users to avoid debating which VAT they trust more to create reliable obfuscation features. This contrasts with our (and prior research [6, 97]) conceptualization of obfuscation techniques as a feature within VAT. Future research should explore the various workflows of having separate privacy techniques outside VAT compared to embedding them within VAT.

## 6 Limitations

We conducted a qualitative study to gather in-depth insights without introducing a technology prototype that could limit participants' imaginations [38, 82]. However, we were only able to capture participants' anticipated use. Extending our findings, future research can build obfuscation techniques with assessment descriptors and study blind people's perspectives. Furthermore, the majority of our participants are totally blind. Exploring the experience of people with diverse visual disabilities may reveal additional considerations. Additionally, our participants did not comment on having additional disabilities. Future work could engage with multiply disabled blind groups to understand their perspectives on obfuscation and assessment descriptors. Finally, all our participants are located in the United States. While our findings may be applicable in other contexts, upcoming research can investigate the potential benefits and drawbacks of assessment descriptors in regions beyond the United States.

## 7 Conclusion

This study explored how to design for accessible error detection in emerging AI-enabled privacy techniques (obfuscation) in VAT. Particularly, we examined the potential benefits and drawbacks of assessment descriptors. Through interviews and focus groups, we found that vague and highly visual assessment descriptors, specifically an object's color, dimensions, and distance from the user, are insufficient in supporting blind people in detecting errors. Alternatively, participants shared other assessment descriptors that better represent how blind people make sense of errors, such as describing multiple objects within a familiar space. Furthermore, participants pointed out that assessment descriptors are a small part of negotiating trust, advocating for transparency on how AI-enabled privacy techniques are created, and co-creating training materials for how to use obfuscation. These findings suggest a need to challenge (sighted) designers' and researchers' underlying assumptions when developing assessment descriptors and rethink emerging AI-enabled techniques more holistically, considering how to develop community-centered onboarding materials and establish support during obfuscation errors.

## Acknowledgments

We thank and appreciate the participants for sharing their perspectives with us, and the National Federation of the Blind (NFB) for circulating our recruitment survey. Additionally, we would like to express our gratitude to Novia Nurian and Aashaka Desai for providing generous feedback on earlier drafts of this paper. We also thank the anonymous reviewers for their constructive comments.

This material is based upon work supported by the National Science Foundation under Grant 1763297, a Google Research Inclusion Award, and a Predoctoral Fellowship from the University of Michigan. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the funding organizations.

## References

[1] Ali Abdolrahmani, William Easley, Michele Williams, Stacy Branham, 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*. 4158–4169.

[2] Rudaiba Adnin and Maitraye Das. 2024. "I look at it as the king of knowledge": How Blind People Use and Understand Generative AI Tools. *people* 16, 54 (2024), 92.

[3] Aira. 2024. Aira. https://aira.io/aira-live

[4] 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.

[5] Taslima Akter, Manohar Swaminathan, and Apu Kapadia. 2024. Toward Effective Communication of AI-Based Decisions in Assistive Tools: Conveying Confidence and Doubt to People with Visual Impairments at Accelerated Speech. In *Proceedings of the 21st International Web for All Conference*. 177–189.

[6] Rahaf Alharbi, Robin N Brewer, and Sarita Schoenebeck. 2022. Understanding emerging obfuscation technologies in visual description services for blind and low vision people. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–33.

[7] 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*. 1–17.

[8] Rawan Alharbi, Tammy Stump, Nilofar Vafaie, Angela Pfammatter, Bonnie Spring, and Nabil Alshurafa. 2018. I can't be myself: effects of wearable cameras on the capture of authentic behavior in the wild. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 2, 3 (2018), 1–40.

[9] Rawan Alharbi, Mariam Tolba, Lucia C Petito, Josiah Hester, and Nabil Alshurafa. 2019. To mask or not to mask? balancing privacy with visual confirmation utility in activity-oriented wearable cameras. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 3, 3 (2019), 1–29.

[10] American Foundation for the Blind. 2022. Apple Releases Impressive Updates in iOS 16 for People Who Are Blind or Low Vision. *AccessWorld* 22, 5 (2022). https://www.afb.org/aw/22/5/17555 Accessed: 2024-09-13.

[11] Arslan Anwar. 2022. OneStep Reader. https://blindhelp.net/software/onestep-reader Accessed: 2025-04-09.

[12] Mauro Avila, Katrin Wolf, Anke Brock, and Niels Henze. 2016. Remote assistance for blind users in daily life: A survey about be my eyes. In *Proceedings of the 9th ACM International Conference on PErvasive Technologies Related to Assistive Environments*. 1–2.

[13] Be My Eyes. 2023. Image to Text AI and Blurred Faces: What's Going On. https://www.bemyeyes.com/blog/image-to-text-ai-and-blurred-faces-whats-going-on. Accessed: 2024-09-30.

[14] Be My Eyes. n.d.. Be My Eyes for Windows. https://www.bemyeyes.com/be-my-eyes-for-windows/ Accessed: 2025-04-13.

[15] Application BeMyEyes. 2024. www.bemyeyes.com

[16] Emily M Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the dangers of stochastic parrots: Can language models be too big?. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*. 610–623.

[17] Cynthia L Bennett, Erin Brady, and Stacy M Branham. 2018. Interdependence as a frame for assistive technology research and design. In *Proceedings of the 20th international acm sigaccess conference on computers and accessibility*. 161–173.

[18] Cynthia L Bennett, Cole Gleason, Morgan Klaus Scheuerman, Jeffrey P Bigham, Anhong Guo, and Alexandra To. 2021. "It's Complicated": Negotiating Accessibility and (Mis) Representation in Image Descriptions of Race, Gender, and Disability. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–19.

[19] Cynthia L Bennett and Daniela K Rosner. 2019. The Promise of Empathy: Design, Disability, and Knowing the" Other". In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–13.

[20] Cynthia L Bennett, Daniela K Rosner, and Alex S Taylor. 2020. The care work of access. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–15.

[21] Patty Berne. 2015. Disability Justice: A Working Draft. https://www.sinsinvalid.org/blog/disability-justice-a-working-draft-by-patty-berne Accessed: 2024-07-02.

[22] Maalvika Bhat and Duri Long. 2024. Designing Interactive Explainable AI Tools for Algorithmic Literacy and Transparency. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference*. 939–957.

[23] Nilavra Bhattacharya, Qing Li, and Danna Gurari. 2019. Why does a visual question have different answers?. In *Proceedings of the IEEE/CVF international conference on computer vision*. 4271–4280.

[24] Elettra Bietti. 2020. From ethics washing to ethics bashing: a view on tech ethics from within moral philosophy. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*. 210–219.

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

[26] 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*. 2117–2126.

[27] Danielle Bragg, Oscar Koller, Naomi Caselli, and William Thies. 2020. Exploring collection of sign language datasets: Privacy, participation, and model performance. In *Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility*. 1–14.

[28] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[29] Virginia Braun and Victoria Clarke. 2019. Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health* 11, 4 (2019), 589–597.

[30] Virginia Braun and Victoria Clarke. 2023. Toward good practice in thematic analysis: Avoiding common problems and be (com) ing a knowing researcher. *International journal of transgender health* 24, 1 (2023), 1–6.

[31] Michele A Burton, Erin Brady, Robin Brewer, Callie Neylan, Jeffrey P Bigham, and Amy Hurst. 2012. Crowdsourcing subjective fashion advice using VizWiz: challenges and opportunities. In *Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility*. 135–142.

[32] Ángel Alexander Cabrera, Adam Perer, and Jason I Hong. 2023. Improving human-AI collaboration with descriptions of AI behavior. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–21.

[33] Ana Caraban, Evangelos Karapanos, Daniel Gonçalves, and Pedro Campos. 2019. 23 ways to nudge: A review of technology-mediated nudging in human-computer interaction. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–15.

[34] Danielle Keats Citron and Daniel J Solove. 2021. Privacy Harms. *Available at SSRN* (2021).

[35] Anna L Cox, Sandy JJ Gould, Marta E Cecchinato, Ioanna Iacovides, and Ian Renfree. 2016. Design frictions for mindful interactions: The case for microboundaries. In *Proceedings of the 2016 CHI conference extended abstracts on human factors in computing systems*. 1389–1397.

[36] Badhan Chandra Das, M Hadi Amini, and Yanzhao Wu. 2025. Security and privacy challenges of large language models: A survey. *Comput. Surveys* 57, 6 (2025), 1–39.

[37] Terrance De Vries, Ishan Misra, Changhan Wang, and Laurens Van der Maaten. 2019. Does object recognition work for everyone?. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*. 52–59.

[38] Nicola Dell, Vidya Vaidyanathan, Indrani Medhi, Edward Cutrell, and William Thies. 2012. " Yours is better!" participant response bias in HCI. In *Proceedings of the sigchi conference on human factors in computing systems*. 1321–1330.

[39] Mariella Dimiccoli, Juan Marín, and Edison Thomaz. 2018. Mitigating bystander privacy concerns in egocentric activity recognition with deep learning and intentional image degradation. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 4 (2018), 1–18.

[40] Pei Du and Nirupama Bulusu. 2021. An automated AR-based annotation tool for indoor navigation for visually impaired people. In *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility*. 1–4.

[41] Upol Ehsan, Q Vera Liao, Michael Muller, Mark O Riedl, and Justin D Weisz. 2021. Expanding explainability: Towards social transparency in ai systems. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. 1–19.

[42] Envision. 2025. Perceive Possibility. https://www.letsenvision.com/ Accessed: 2025-04-15.

[43] Envision. n.d.. Envision Glasses: Home Edition. https://www.letsenvision.com/glasses/home Accessed: 2025-04-13.

[44] Alexander Erlei, Abhinav Sharma, and Ujwal Gadiraju. 2024. Understanding choice independence and error types in human-ai collaboration. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–19.

[45] Sandra Fernando, Chiemela Ndukwe, Bal Virdee, and Ramzi Djemai. 2025. Image recognition tools for blind and visually impaired users: An emphasis on the design considerations. *ACM Transactions on Accessible Computing* 18, 1 (2025), 1–21.

[46] Freedom Scientific. n.d.. New and Improved Features in JAWS. https://www.freedomscientific.com/training/jaws/new-and-improved-features/ Accessed: 2025-04-13.

[47] 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*. 1–21.

[48] Patricia Garcia, Tonia Sutherland, Marika Cifor, Anita Say Chan, Lauren Klein, Catherine D'Ignazio, and Niloufar Salehi. 2020. No: Critical refusal as feminist data practice. In *Companion Publication of the 2020 Conference on Computer Supported Cooperative Work and Social Computing*. 199–202.

[49] Bhavya Ghai, Q Vera Liao, Yunfeng Zhang, Rachel Bellamy, and Klaus Mueller. 2021. Explainable active learning (xal) toward ai explanations as interfaces for machine teachers. *Proceedings of the ACM on human-computer interaction* 4, CSCW3 (2021), 1–28.

[50] Kate S Glazko, Momona Yamagami, Aashaka Desai, Kelly Avery Mack, Venkatesh Potluri, Xuhai Xu, and Jennifer Mankoff. 2023. An Autoethnographic Case Study of Generative Artificial Intelligence's Utility for Accessibility. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–8.

[51] Ricardo E Gonzalez Penuela, Jazmin Collins, Cynthia Bennett, and Shiri Azenkot. 2024. Investigating Use Cases of AI-Powered Scene Description Applications for Blind and Low Vision People. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–21.

[52] Danna Gurari, Qing Li, Chi Lin, Yinan Zhao, Anhong Guo, Abigale Stangl, and Jeffrey P Bigham. 2019. Vizwiz-priv: A dataset for recognizing the presence and purpose of private visual information in images taken by blind people. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 939–948.

[53] 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*. 3608–3617.

[54] Michael Hansen. 2014. OneStep Reader. https://www.applevis.com/comment/43380 Accessed: 2025-04-09.

[55] Rakibul Hasan, Eman Hassan, Yifang Li, Kelly Caine, David J Crandall, Roberto Hoyle, and Apu Kapadia. 2018. Viewer experience of obscuring scene elements in photos to enhance privacy. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13.

[56] Jaylin Herskovitz, Andi Xu, Rahaf Alharbi, and Anhong Guo. 2023. Hacking, switching, combining: understanding and supporting DIY assistive technology design by blind people. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–17.

[57] Michael Townsen Hicks, James Humphries, and Joe Slater. 2024. ChatGPT is bullshit. *Ethics and Information Technology* 26, 2 (2024), 1–10.

[58] Jonggi Hong. 2021. *Exploring Blind and Sighted Users' Interactions with Error-Prone Speech and Image Recognition*. Ph. D. Dissertation. University of Maryland, College Park.

[59] Jonggi Hong, Jaina Gandhi, Ernest Essuah Mensah, Farnaz Zamiri Zeraati, Ebrima Jarjue, Kyungjun Lee, and Hernisa Kacorri. 2022. Blind Users Accessing Their Training Images in Teachable Object Recognizers. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–18.

[60] Jonggi Hong and Hernisa Kacorri. 2024. Understanding How Blind Users Handle Object Recognition Errors: Strategies and Challenges. *arXiv preprint arXiv:2408.03303* (2024).

[61] Noura Howell, Watson F Hartsoe, Jacob Amin, and Vyshnavi Namani. 2024. Reflective Design for Informal Participatory Algorithm Auditing: A Case Study with Emotion AI. In *Proceedings of the 13th Nordic Conference on Human-Computer Interaction*. 1–17.

[62] Janet Ingber. 2018. Envision AI and Seeing AI: Two Multi-Purpose Recognition Apps. *AccessWorld* 19, 9 (2018). https://www.afb.org/aw/19/9/15059

[63] IrisVision. n.d.. IrisVision. https://irisvision.com/ Accessed: 2025-04-13.

[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*. 203–210.

[65] Chutian Jiang, Emily Kuang, and Mingming Fan. 2025. How Can Haptic Feedback Assist People with Blind and Low Vision (BLV): A Systematic Literature Review. *ACM Transactions on Accessible Computing* 18, 1 (2025), 1–57.

[66] Harmanpreet Kaur, Matthew R Conrad, Davis Rule, Cliff Lampe, and Eric Gilbert. 2024. Interpretability Gone Bad: The Role of Bounded Rationality in How Practitioners Understand Machine Learning. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW1 (2024), 1–34.

[67] Sunnie SY Kim, Elizabeth Anne Watkins, Olga Russakovsky, Ruth Fong, and Andrés Monroy-Hernández. 2023. Humans, ai, and context: Understanding end-users' trust in a real-world computer vision application. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. 77–88.

[68] Tae-hoon Kim, Dongmin Kang, Kari Pulli, and Jonghyun Choi. 2019. Training with the invisibles: Obfuscating images to share safely for learning visual recognition models. *arXiv preprint arXiv:1901.00098* (2019).

[69] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. 2023. Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 4015–4026.

[70] Hao-Ping Lee, Yu-Ju Yang, Thomas Serban Von Davier, Jodi Forlizzi, and Sauvik Das. 2024. Deepfakes, phrenology, surveillance, and more! a taxonomy of ai privacy risks. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–19.

[71] Kyungjun Lee, Jonggi Hong, Simone Pimento, Ebrima Jarjue, and Hernisa Kacorri. 2019. Revisiting blind photography in the context of teachable object recognizers. In *Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility*. 83–95.

[72] Franklin Mingzhe Li, Franchesca Spektor, Meng Xia, Mina Huh, Peter Cederberg, Yuqi Gong, Kristen Shinohara, and Patrick Carrington. 2022. "It feels like taking a gamble": Exploring perceptions, practices, and challenges of using makeup and cosmetics for people with visual impairments. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–15.

[73] Zongxia Li, Xiyang Wu, Hongyang Du, Huy Nghiem, and Guangyao Shi. 2025. Benchmark evaluations, applications, and challenges of large vision language models: A survey. *arXiv preprint arXiv:2501.02189* 1 (2025).

[74] Haley MacLeod, Cynthia L Bennett, Meredith Ringel Morris, and Edward Cutrell. 2017. Understanding blind people's experiences with computer-generated captions of social media images. In *proceedings of the 2017 CHI conference on human factors in computing systems*. 5988–5999.

[75] Atefeh Mahdavi Goloujeh, Anne Sullivan, and Brian Magerko. 2024. Is It AI or Is It Me? Understanding Users' Prompt Journey with Text-to-Image Generative AI Tools. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–13.

[76] 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*. 3–10.

[77] Daniela Massiceti, Camilla Longden, Agnieszka Slowik, Samuel Wills, Martin Grayson, and Cecily Morrison. 2023. Explaining CLIP's performance disparities on data from blind/low vision users. *arXiv preprint arXiv:2311.17315* (2023).

[78] 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*. 1–5.

[79] Cormac McGrath, Per J Palmgren, and Matilda Liljedahl. 2019. Twelve tips for conducting qualitative research interviews. *Medical teacher* 41, 9 (2019), 1002–1006.

[80] Thomas Mejtoft, Emma Parsjö, Ole Norberg, and Ulrik Söderström. 2023. Design Friction and Digital Nudging: Impact on the Human Decision-Making Process. In *Proceedings of the 2023 5th International Conference on Image, Video and Signal Processing*. 183–190.

[81] Rod Michalko. 2001. Blindness enters the classroom. *Disability & Society* 16, 3 (2001), 349–359.

[82] Joy Ming, Sharon Heung, Shiri Azenkot, and Aditya Vashistha. 2021. Accept or address? Researchers' perspectives on response bias in accessibility research. In *Proceedings of the 23rd International ACM SIGACCESS conference on computers and accessibility*. 1–13.

[83] Kyle Montague. 2010. Accessible indoor navigation. In *Proceedings of the 12th international ACM SIGACCESS conference on Computers and accessibility*. 305–306.

[84] Kyzyl Monteiro, Yuchen Wu, and Sauvik Das. 2024. Manipulate to Obfuscate: A Privacy-Focused Intelligent Image Manipulation Tool for End-Users. In *Adjunct Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–3.

[85] Meredith Ringel Morris, Jazette Johnson, Cynthia L Bennett, and Edward Cutrell. 2018. Rich representations of visual content for screen reader users. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–11.

[86] Cecily Morrison, Martin Grayson, 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*. 1–12.

[87] Amal Nanavati, Xiang Zhi Tan, and Aaron Steinfeld. 2018. Coupled indoor navigation for people who are blind. In *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. 201–202.

[88] Leo Neat, Ren Peng, Siyang Qin, and Roberto Manduchi. 2019. Scene text access: A comparison of mobile OCR modalities for blind users. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*. 197–207.

[89] Michele Paris. 2023. Introducing Be My AI. https://www.bemyeyes.com/blog/introducing-be-my-ai. Accessed: April 7, 2024.

[90] Hana Porkertová. 2022. Revising modern divisions between blindness and sightedness: Doing knowledge in blind assemblages. *The Sociological Review* 70, 3 (2022), 580–598.

[91] Silvia Rodríguez Vázquez. 2016. Measuring the impact of automated evaluation tools on alternative text quality: a web translation study. In *Proceedings of the 13th International Web for All Conference*. 1–10.

[92] Siegfried Saerberg. 2010. "Just go straight ahead" how blind and sighted pedestrians negotiate space. *The Senses and Society* 5, 3 (2010), 364–381.

[93] Freedom Scientific. n.d.. Speech and Sounds Schemes. https://support.freedomscientific.com/SurfsUp/18-SpeechSoundsSchemes.htm Accessed: 2024-09-11.
[94] SeeingAI. 2024. SeeingAI. https://www.microsoft.com/en-us/ai/seeing-ai
[95] Jonathan Silvertown. 2009. A new dawn for citizen science. *Trends in ecology & evolution* 24, 9 (2009), 467–471.
[96] Abigale Stangl, Meredith Ringel Morris, and Danna Gurari. 2020. " Person, Shoes, Tree. Is the Person Naked?" What People with Vision Impairments Want in Image Descriptions. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.
[97] Abigale Stangl, Emma Sadjo, Pardis Emami-Naeini, Yang Wang, Danna Gurari, and Leah Findlater. 2023. "Dump it, Destroy it, Send it to Data Heaven": Blind People's Expectations for Visual Privacy in Visual Assistance Technologies. In *Proceedings of the 20th International Web for All Conference*. 134–147.
[98] Abigale Stangl, Kristina Shiroma, Nathan Davis, Bo Xie, Kenneth R Fleischmann, Leah Findlater, and Danna Gurari. 2022. Privacy concerns for visual assistance technologies. *ACM Transactions on Accessible Computing (TACCESS)* 15, 2 (2022), 1–43.
[99] Abigale Stangl, Kristina Shiroma, Bo Xie, Kenneth R Fleischmann, and Danna Gurari. 2020. Visual Content Considered Private by People Who are Blind. In *The 22nd International ACM SIGACCESS Conference on Computers and Accessibility*. 1–12.
[100] Cella M Sum, Rahaf Alharbi, Franchesca Spektor, Cynthia L Bennett, Christina N Harrington, Katta Spiel, and Rua Mae Williams. 2022. Dreaming disability justice in HCI. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 1–5.
[101] TechCrunch. 2020. iPhones can now tell blind users where and how far away people are. *TechCrunch* (2020). https://techcrunch.com/2020/10/30/iphones-can-now-tell-blind-users-where-and-how-far-away-people-are/ Accessed: 2024-09-13.
[102] The Program on Intergroup Relations. 2023. *Developing Community Guidelines*. University of Michigan. https://igr.umich.edu/research-publications/igr-insight IGR Insight handout.
[103] Tanya Titchkosky, Devon Healey, and Rod Michalko. 2019. Blindness simulation and the culture of sight. *Journal of Literary & Cultural Disability Studies* 13, 2 (2019), 123–139.
[104] Violet Turri, Katelyn Morrison, Katherine-Marie Robinson, Collin Abidi, Adam Perer, Jodi Forlizzi, and Rachel Dzombak. 2024. Transparency in the Wild: Navigating Transparency in a Deployed AI System to Broaden Need-Finding Approaches. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*. 1494–1514.
[105] TWIML AI Podcast. 2021. Accessibility and Computer Vision. https://twimlai.com/podcast/twimlai/accessibility-computer-vision/. Accessed: 2024-09-30.
[106] Beatrice Vincenzi, Alex S Taylor, and Simone Stumpf. 2021. Interdependence in Action: People with Visual Impairments and their Guides Co-constituting Common Spaces. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–33.
[107] W3C Web Accessibility Initiative. 2016. Images Tutorial: Complex Images. https://www.w3.org/WAI/tutorials/images/complex/ Accessed: 2025-04-07.
[108] W3C Web Accessibility Initiative (WAI). 2021. Images Tutorial – Web Accessibility Tutorials. https://www.w3.org/WAI/tutorials/images/ Accessed: 2025-04-14.
[109] Jonas Wachner, Marieke Adriaanse, and Denise De Ridder. 2021. The influence of nudge transparency on the experience of autonomy. *Comprehensive Results in Social Psychology* 5, 1-3 (2021), 49–63.
[110] Yang Wang and Charlotte Emily Price. 2022. Accessible privacy. In *Modern socio-technical perspectives on privacy*. Springer International Publishing Cham, 293–313.
[111] Ben Whitburn and Rod Michalko. 2019. Blindness/sightedness: Disability studies and the defiance of di-vision. In *Routledge handbook of disability studies*. Routledge, 219–233.
[112] Michele A Williams, Caroline Galbraith, Shaun K Kane, and Amy Hurst. 2014. " just let the cane hit it" how the blind and sighted see navigation differently. In *Proceedings of the 16th international ACM SIGACCESS conference on Computers & accessibility*. 217–224.
[113] Rua Mae Williams, Louanne Boyd, and Juan E Gilbert. 2023. Counterventions: a reparative reflection on interventionist HCI. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–11.
[114] Rua M Williams and LouAnne E Boyd. 2019. Prefigurative politics and passionate witnessing. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*. 262–266.
[115] Monika Zalnieriute. 2021. " Transparency Washing" in the Digital Age: A Corporate Agenda of Procedural Fetishism. *Critical Analysis L.* 8 (2021), 139.
[116] Angie Zhang, Rocita Rana, Alexander Boltz, Veena Dubal, and Min Kyung Lee. 2024. Data Probes as Boundary Objects for Technology Policy Design: Demystifying Technology for Policymakers and Aligning Stakeholder Objectives in Rideshare Gig Work. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–21.
[117] Jingyi Zhang, Jiaxing Huang, Sheng Jin, and Shijian Lu. 2024. Vision-language models for vision tasks: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2024).
[118] Lotus Zhang, Abigale Stangl, Tanusree Sharma, Yu-Yun Tseng, Inan Xu, Danna Gurari, Yang Wang, and Leah Findlater. 2024. Designing Accessible Obfuscation Support for Blind Individuals' Visual Privacy Management. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–19.
[119] Zhuohao Jerry Zhang, Smirity Kaushik, JooYoung Seo, Haolin Yuan, Sauvik Das, Leah Findlater, Danna Gurari, Abigale Stangl, and Yang Wang. 2023. {ImageAlly}: A {Human-AI} Hybrid Approach to Support Blind People in Detecting and Redacting Private Image Content. In *Nineteenth Symposium on Usable Privacy and Security (SOUPS 2023)*. 417–436.
[120] Yuhang Zhao, Shaomei Wu, Lindsay Reynolds, and Shiri Azenkot. 2017. The effect of computer-generated descriptions on photo-sharing experiences of people with visual impairments. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 1–22.

# A Describing Focus and Background Modes to Participants During Interviews and Focus Groups

To describe focus mode, we provided the following definition and example:

> Imagine that [a VAT that the participants use frequently] added a new feature that aims to preserve your privacy. It is called "focus mode," highlighting a specific object you're interested in, while a blur filter would hide everything else. For example, say you are standing in the kitchen and want to use [VAT] to find out your milk's expiration date. If you use focus mode, the camera will only show milk, and anything else, like your kitchen counters, will be blurred and not be shown.

We defined background mode as:

> Option two is called "background mode," which hides specific elements you decided on beforehand. For example, you may choose pill bottles as private content and turn on the background mode feature. This would detect if there is a pill bottle in the background, and it would blur or hide it while showing everything else.

The examples following background and focus modes were inspired by cases presented in prior research [6]. However, we asked participants to share with us other examples that are relevant to their everyday use of VAT.

# B Interview Questions

**NOTE:** We followed a semi-structured approach. The questions we listed in this section are merely guidelines. In the interviews, we refined these questions and asked follow-up questions. This is also a *segment* of our full interview protocol that is relevant to this paper.

After describing focus mode (refer to Appendix A), we asked participants:

## B.1 Focus Mode Thoughts & Reactions

- What do you think of the focus mode feature?
- When would you use this focus mode feature? Why?
- When would you NOT use this focus mode feature? Why?
- What are some benefits that you might gain from using this focus mode feature? (probe: ask participants to also think

about societal benefits or benefits to the blind community at large).
- What are some of the harms or risks associated with this hypothetical feature? (probe: ask participants to also think about societal risks or risks to the blind community at large).

## B.2 Understanding Focus Mode

- Generally, how would you imagine assessing the quality of this focus mode feature?
- Would your process of evaluating the quality of the focus mode feature differ from how you assess the quality of VAT in general?
- How would you like to know if the focus mode is not working as expected?
- How would you like to know if the focus mode is working correctly without any errors?
- Do you think there might be potential risks associated with [insert participant suggestions]?
- How do you imagine VAT could address or decrease this risk?
- How would you imagine this to work differently if you were using a human-enabled VAT? (probe: assessing quality, benefits, harms, and risks)

After describing background mode (refer to Appendix A) we asked participants:

## B.3 Background Mode Thoughts & Reactions

- What do you think of the background mode?
- What are some benefits that you might gain from using this mode? (probe: ask participants to also think about societal benefits or benefits to the blind community at large).
- What are some of the harms or risks associated with this hypothetical feature? (probe: ask participants to also think about societal risks or risks to blind community at large).
- Beyond the pill bottle example I gave, when would you use this feature? Why?
- When would you **NOT** use this background mode feature? why?

## B.4 Understanding Background Mode

- In previous sections of this interview, we talked about confidence and quality when using VAT. Generally, how would you imagine assessing the quality of this background mode feature?
- Would your process of evaluating the quality of the background mode feature differ from how you assess the quality of VAT in general?
- How would you like to know if the background mode is not working as expected?
- How would you like to know if the background mode is working correctly without any errors?
- Do you think there might be potential risks associated with [insert participant suggestions]?
- How do you imagine VAT could address or decrease this risk?
- How would you imagine this to work differently if you were using a human-enabled VAT? (probe: assessing quality, benefits, harms, and risks)

## B.5 Focus Mode Vs. Background Mode

- In general, do you prefer background mode or focus mode? Why?
- In what situations would you prefer using focus mode over background mode? Why?
- In what situations would you prefer using background mode over focus mode? Why?
- Would you say one mode has more risks or trade offs than the other? If so, which and why?

# C Focus Groups Questions

**NOTE:** We followed a semi-structured approach. The questions we listed in this section are merely guidelines. In the focus groups, we refined these questions and asked follow-up questions.

## C.1 Introduction and Welcome Activity

Hi everyone, my name is [researcher name]. I use she/her pronouns.

Thank you everyone for being here today. Our conversation is going to last for about an hour and a half. It could end earlier, but it won't go longer.

Today, we will brainstorm ways to make visual assistance technologies like Seeing AI, Aira, and Be My Eyes include better privacy and control features that are accessible to blind and low vision people. Before we get into the research details, let's do an icebreaker to get to know each other:

- Can everyone say their first names or an alias you want to be called by during today's session? And optionally, you could share your pronouns.
- *Group norms:*
  - Does anyone have preferences or ground rules they would like to share for group conversations?
  - I also want to establish some group expectations:
    - Don't share any info spoken here outside the group.
    - I encourage you all to talk to each other, ask questions, and comment on each other's thoughts and points of view.
    - In our focus group, we're not aiming for collective agreement. Embracing differences and *respectful* disagreements are welcomed.
    - Challenge the idea and not the person. If we wish to challenge something that has been said, we will challenge the idea or the practice referred to, not the individual sharing this idea or practice.
    - At times, we can also "agree to disagree," so don't feel pressured to agree just because others might be leaning a certain way.
    - If you tend to speak more, make sure to leave airtime for others to share.
    - If you tend to be quieter, challenge yourself to contribute! We would love to hear and learn from you.

    * I might call on participants to answer questions, but you are always welcome to skip any question that you don't want to answer.
    * Does anyone have additional expectations they would like to share?

## C.2 Overview of Research Activities

Today, we will go through a few activities to explore two hypothetical features for visual assistance technologies. I'll share some fictional scenarios just as a way to spark our discussion.

As you may remember, option one is called "focus mode." If a user turns on this feature and selects a type of object to highlight, the user's camera would only view and get access to that specific object, while everything else would be hidden by blurring or masking. Option two is called "background mode." If a user turns on this feature and selects a type of object to hide, it would blur or mask that specific type of object while preserving everything else around it.

So every time you hear focus mode, think of one object as visible and everything else is blurred. Every time you hear background mode, think of the background as visible, and one object is blurred. These filters are just thought of as a first layer for you to get more control and increase before moving on to other features like short text, document mode, scene preview, or features in other applications.

From our interviews, we learned that a key challenge with these features is how blind and low vision users know if the features are working properly without relying on sight. Today, we're eager to hear your ideas and suggestions for how we can ensure these features will be easy to use and trustworthy for blind people. Your thoughts are valuable, and I am looking forward to hearing them!

## C.3 Audio Probe 1

For the first activity, I'm going to play an audio recording of this fictional user scenario. Before I play the audio, I just want to note that the audio is deliberately short and is missing some context, so we can build the story together! Sometimes the object or task might not be meaningful. I am hoping we can focus on the confirmation message that focus mode gives the feature first, and then we can discuss how this might be helpful or frustrating in other scenarios.

After playing audio probe 1 (refer to Appendix D), I ask the following questions:

- How do you feel about hearing the dimensions information of the boxed meal when you use the focus mode feature?
  - How might knowing the object's dimensions affect how much you trust focus mode?
  - In addition to the dimensions of the microwavable meal, what other information would help you to understand how well focus mode is working?
  - Instead of the dimensions of the microwavable meal, how about knowing how much the microwavable meal is visible within the camera frame?
- How can this information on the boxed meal's dimensions be communicated in a way that is more intuitive for you?
  - If the dimensions of microwavable meals were conveyed through a sensory component, such sound, how would you prefer this information to be presented?
  - If the size of microwavable meals had a sound, what types of sounds would make it easier for you to understand the size? Can you think of tones, pitches, or rhythms that would represent different dimensions?
  - How about touch? What are the ways touch can be used to communicate the box dimensions and how much of the image is visible and how much is hidden?
- In this scenario, if focus mode was included in Seeing AI or TapTapSee, what are your concerns about using the focus mode feature?
  - What factors would make you skeptical about the provided measurements?
- So thinking back to a recent experience of you using a visual assistance application and potentially wanting to use focus mode. Walk me through this example and the kind of information you would like to know to get a better sense of how focus mode is working?
- How might having additional information about the object impact your experience when using human-based applications like Aira or Be My Eyes volunteers?
- What are the different concerns that might arise when using human-based applications like Aira agents or Be My Eyes volunteers?

## C.4 Audio Probe 2

Thank you for sharing your perspectives so far. I am going to play another audio clip now. Before I play the audio, I just want to note that this audio clip is also deliberately short and is missing some context, so we can build the story together!

After playing audio probe 2 (refer to Appendix D), I ask the following questions:

- Does having specific details about what it's hiding, specifically the box's approximate location and color, help you better understand the background mode feature?
- How do details, like color and distance after hiding an object, affect your trust in using background mode?
- When using background mode, can you think of other situations where descriptions of the object (in this case, color and location) may cause confusion or frustration?
- As you recall, background mode described the object of interest, a delivery box, as "A brown box about 3 ft away from you is blurred while everything else is visible." What are ways that this message can be improved?
  - What changes would you suggest to make the message more clear and easier to understand? How do you imagine alternative ways this message would feel or sound like?
  - Besides color and distance, what other info would be helpful for you to understand if background mode is working properly?
- In this scenario, if background mode was included in Seeing AI or TapTapSee, what are your concerns about using the focus mode feature?

- So thinking back to a recent experience of you using a visual assistance application and potentially wanting to use background mode. Walk me through this example and the kind of information you would like to know to get a better sense of how background mode is working?
- Let's say you are at a dinner party and you wanted to take photos of the food without including the guests. You turned on background mode and typed faces. What is the confirmation message you want to get from background mode?
- To recap, background mode would give you descriptive messages of the color, distance, or other details of the hidden object to let you know that the object has been blurred. How might background mode's descriptive messages impact how you use it with a human-based VAT such as Be My Eyes or Aira?

## C.5 Before and After Using Obfuscation Techniques

Now, let's transition to talking about using background mode and focus mode in your everyday life. So let's move beyond the scenarios that I mentioned previously. Before we get started, does anyone have questions or would like me to repeat anything?

### *C.5.1 Before Using Obfuscation Techniques:*

- If you were using focus mode and background mode for the first time ever, what questions would you have?
- What key details would you like to know upfront about how 'focus mode' and 'background mode' work? Why do you think these details are important?
- If you were among the team of people who are creating and advertising focus mode and background mode, how would you ensure that users can easily find the limitations of 'focus mode' and 'background mode'?
- How often would you like to hear updates about the limits and concerns of 'focus mode' and 'background mode'? Why?

### *C.5.2 After Using Obfuscation:*

- Think about a recent time when you gave feedback to a person. It could be a family member, a friend, or a colleague. Walk me through the actions or responses from them that made you feel truly heard? (If you haven't given feedback to a person before, how would you imagine feeling heard? What kind of response or action from them do you think would make you feel like your input matters?)
- Now, imagine giving feedback to an application or website, what would make you feel acknowledged and valued? Is there a particular way you'd hope the application or website would respond to your input?
- If you had concerns or questions about the outcomes of using the 'background mode" or "focus mode" features, what are some ways you might want to reach out to the visual assistance company for support or clarification?
- If you encounter a problem with the focus mode or background mode, what kind of response would you expect from the application?
  - What would the ideal next steps be?
  - How involved would you want to be in that process?
  - Imagine a human customer service agent trying to fix accuracy issues with these features (background and focus mode).
    - How might you describe the issue to them?
    - How likely would you be to trust them to fix the issue?
  - Now, let's imagine the customer service agent was a chatbot.
    - How might you describe the issue to the chatbot?
    - How would you expect the chatbot to interact to fix this issue?
    - How likely would you be to trust the chatbot to fix the issue?
- In the event of a concern with 'background mode' or 'focus mode,' how important is it for you that the visual assistance company provides the option to delete your personal data associated with these visual assistance technologies?

## C.6 Closing Remarks

- Based on our discussion, would anyone like to add something we missed?
- Does anyone have any questions for me?

# D Scripts of Audio Probes

**Table 2: Scripts of audio probes played in focus groups. The hypothetical user was a different voice actor than the interviewers. We used text-to-speech software (TTS) to simulate the interaction.**

| Focus Mode Audio Probe | Background Mode Audio Probe |
|---|---|
| *User:* OK, what's the cooking instructions for this microwavable meal? Let me use an AI app like Seeing AI or TapTapSee... But wait, let me turn on Focus mode.<br>`TTS: Focus Mode. What would you like to highlight?`<br>*User:* OK I'm going to type microwavable meals.<br>`TTS: An object that is approximately 8 inches in length and 4 inches in width is highlighted while everything else is blurred.`<br>*User:* Cool! Let's get those cooking instructions! | *User:* Okay, I need to use a visual assistance application, but I want to make sure that this delivery box that has my name and address is not visible. Let me turn on background mode.<br>`TTS: Background mode. What do you want to hide?`<br>*User:* Ok, let me type the delivery box.<br>`TTS: A brown box that is about 3 ft away from you is hidden while everything else is visible.`<br>*User:* Great! |