Misfitting With AI: How Blind People Verify and Contest AI Errors

Rahaf Alharbi
rmalharb@umich.edu
University of Michigan
Ann Arbor, USA

Pa Lor
palor@umich.edu
University of Michigan
Ann Arbor, USA

Jaylin Herskovitz
jayhersk@umich.edu
University of Michigan
Ann Arbor, USA

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

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

Abstract

Blind people use artificial intelligence-enabled visual assistance technologies (AI VAT) to gain visual access in their everyday lives, but these technologies are embedded with errors that may be difficult to verify non-visually. Previous studies have primarily explored sighted users' understanding of AI output and created vision-dependent explainable AI (XAI) features. We extend this body of literature by conducting an in-depth qualitative study with 26 blind people to understand their verification experiences and preferences. We begin by describing errors blind people encounter, highlighting how AI VAT fails to support complex document layouts, diverse languages, and cultural artifacts. We then illuminate how blind people make sense of AI through experimenting with AI VAT, employing non-visual skills, strategically including sighted people, and cross-referencing with other devices. Participants provided detailed opportunities for designing accessible XAI, such as affordances to support contestation. Informed by disability studies framework of misfitting and fitting, we unpacked harmful assumptions with AI VAT, underscoring the importance of celebrating disabled ways of knowing. Lastly, we offer practical takeaways for Responsible AI practice to push the field of accessible XAI forward.

CCS Concepts

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

Keywords

Accessibility, Visual Assistance Technology, Blind people, Verification, Artificial Intelligence, Explainability; Seeing AI; Be My Eyes

ACM Reference Format:
Rahaf Alharbi, Pa Lor, Jaylin Herskovitz, Sarita Schoenebeck, and Robin Brewer. 2024. Misfitting With AI: How Blind People Verify and Contest AI Errors. In *The 26th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '24), October 27–30, 2024, St. John's, NL, Canada.* ACM, New York, NY, USA, 17 pages. https://doi.org/10.1145/3663548.3675659

ASSETS '24, October 27–30, 2024, St. John's, NL, Canada

ACM ISBN 979-8-4007-0677-6/24/10
https://doi.org/10.1145/3663548.3675659

1 Introduction

Artificial intelligence (AI) is pervasive in our everyday lives and routines. Computer vision, a type of AI, is integrated into various real-world systems, from social media filters to traffic cameras to medical imaging. Being able to understand, verify, or even contest AI systems has become consequential for myriad activities, ranging from routine interactions to high-stakes decision-making. For example, when posting images of people on Twitter (now X), users visually noticed that the cropping algorithm would only highlight White people, revealing a kind of racial bias [59, 109]. This incident, along with many other cases of users detecting AI limitations and harms [13, 56, 102], motivated a wealth of prior research to study how users make sense of AI [38, 72, 111]. It also pushed Responsible AI researchers, regulators, and activists to advocate for explainable AI (XAI) as a means to support users' decision-making [9, 79, 82, 95].

However, understanding computer vision outputs often relies on a person's visual ability [8, 72, 119]. As a result, much of the literature on XAI has focused on *sighted* experiences and has designed vision-centered XAI measures such as displaying heatmaps to users [72, 83, 114]. Very little is known about how blind people non-visually make sense of AI outputs, and what types of XAI features or otherwise would be useful [63, 81]. This is especially crucial since error detection could be inaccessible to blind people [7, 50, 81], leading to the overtrust [81] or abandonment [73, 97] of AI systems. Thus, we pose and explore the following research questions: how do blind people verify AI results, and what (if any) types of features may support accessible verification?

We focused on AI-enabled visual assistance technologies (AI VAT), which are real-world mobile applications that blind people use daily to gain visual access, from reading to cooking to describing scenes [50, 60, 98, 117]. In essence, we chose AI VAT to capture blind people's *everyday* experience given its ubiquity in blind communities. Furthermore, blind people are concerned about the lack of accuracy in AI VAT, and the accompanying risks of wrong outputs [2, 50, 71]. For instance, when using AI VAT to navigate an unfamiliar space, blind people worry about stigmatizing errors such as mislabeled bathroom signs [2]. While certain XAI features (e.g., conveying confidence ratings [81]) may support blind people in negotiating errors, these features have not been included in commercially available AI VAT.

We conducted semi-structured interviews with 26 blind people who frequently use AI VAT. Our analysis revealed common

errors blind people face when using AI VAT, highlighting processing limitations and cross-cultural bias. Then, we uncovered how blind people verify AI VAT by employing various methods such as: testing AI VAT in low-risk and familiar contexts, engaging non-visual sensemaking skills, strategically including sighted people, and switching between various applications and devices. Lastly, participants outlined opportunities for accessible XAI to support their verification work. They emphasized improved camera guidance within AI VAT, and they had mixed opinions about the added value of confidence ratings. Whether blind people perceived confidence ratings as positive or negative hinged upon the nature of how they used AI VAT. While confidence ratings were sometimes perceived as a way to save time by allowing people to quickly disregard low-accuracy results, other participants felt that the effort it would take to parse this additional information would interrupt their workflow. Participants also desired ways to challenge and refine AI VAT results through direct feedback.

Informed by feminist disability studies [45, 46, 112], we used the concepts of misfitting and fitting [47] to interpret our findings. Rosemarie Garland-Thomson developed the misfitting framework to describe the mutual and ever-shifting entanglements between bodies and environments. Misfitting and fitting "denote an encounter in which two things come together in either harmony or disjunction" [47]. In the context of technology, AI systems are often built for dominant groups, casting those who fall outside of these groups as misfits. We argue that blind people "misfit" in computer vision systems as those technologies are predominately built for sighted people, using sighted data [54, 85]. Blind people who hold marginalized identities experience an additional layer of misfitting when AI VAT fails to attend to their languages and cultural artifacts. Misfitting as an analytical lens positions disabled people as knowers and experts through their lived experience of navigating access. We highlight blind people's epistemic work of verifying AI VAT, calling for future technologies to celebrate and incorporate disabled ways of knowing.

This study has three main contributions. First, we present empirical data on the everyday experience of blind people who use AI VAT. We add to the growing body of literature on AI for visual access [7, 18, 44, 50, 60, 117, 127] by outlining errors faced by blind people, illuminating how they verify AI outcomes, and their preferred roles in improving AI workflows. Our analysis contributes a holistic understanding of blind people's embodied and continuous process of interpreting AI and centers blind people's expectations for upcoming features that could support their existing work. Second, we extend the framework of misfitting and fitting [47] in the context of assistive technologies to unearth the limitations of AI VAT. Third, we offer specific directions for Responsible AI grounded in participants' desires for contestable and accessible XAI.

2 Related Work

We begin by reviewing the intersection of feminist disability studies and assistive technology research, a grounding sensibility in our study. Then, we highlight prior work on visual assistance technologies, focusing on their affordances and limitations. Lastly, we synthesize the emerging area of XAI and its focus on understanding how end-users interpret AI, noting opportunities for accessibility research.

2.1 Feminist Disability Studies & Accessibility

Disability studies is an interdisciplinary and multidisciplinary field that broadly aims to examine the historical, political, and cultural contexts of disability [14, 40, 91]. Recently, disability studies have become a central sensibility to many HCI and accessibility works (e.g., [7, 62, 84, 86]). In their foundational paper, Mankoff et al. argued that disability studies serve as a critical lens to complicate assistive technology research and identify harms, calling on researchers to engage more deeply with disabled people and move away from medical perspectives [84]. In 1994, disability studies scholar Rosemarie Garland-Thomson coined the term *feminist disability studies* [45], emphasizing that disability is integral to feminist scholarship [46]. Contrary to common misconceptions, feminist disability studies do not merely concern the experiences of disabled women. Rather, it provides an analytical framework to examine structures that construct disability [46], attending to questions of fluidity [112] and knowledge making [47, 57].

Accessibility researchers similarly began incorporating feminist disability studies. For instance, Hofmann et al. built upon Mankoff et al. [84] and feminist stances [57, 121] to advocate for 1) acknowledging and challenging ableism[1], 2) recognizing the complexity of disability, and 3) grounding accessibility research in critical disability studies [62]. Bennett et al. drew from feminist disability studies and science, technology, and society (STS) scholarship to attend to the care work of access, reorienting AI assistive technology research from an individualistic focus (i.e., what can a disabled person do) to a more collective approach [18]. Grounded in crip technoscience – a feminist STS and crip theory approach developed by Hamraie and Fritsch to center the expertise of disabled people in the design process [55] – Hsueh et al. reimagined accessible data visualizations by affirming how access is a continual and transformative process [64].

We contribute to the growing area of feminist disability studies in accessibility by applying Rosemarie Garland-Thomson's framework of misfitting & fitting [47] to AI-based assistive technology. Situated in conversations with feminist scholarship on matters of care and dependency [42, 74], misfitting and fitting offers a materialist extension of the social model of disability [107, 115] by directing attention to dynamic and mutually constructive relationships between bodies and worlds. Garland-Thomson explained "the degree to which that shared material world sustains the particularities of our embodied life at any given moment or place determines our fit or misfit" [47]. Misfitting engenders discomfort and incompatibility whereas fitting entails comfort and harmony. People who fit are typically considered "uniform, standard, majority bodies" [47]. For Garland-Thomson, misfitting inspires expertise or crip wisdom [64, 126, 130, 131], highlighting the ways in which disabled people

[1]Talila A. Lewis, community lawyer and organizer, defines ableism as "[a] system of assigning value to people's bodies and minds based on societally constructed ideas of normalcy, productivity, desirability, intelligence, excellence, and fitness. These constructed ideas are deeply rooted in eugenics, anti-Blackness, misogyny, colonialism, imperialism, and capitalism. This systemic oppression that leads to people and society determining people's value based on their culture, age, language, appearance, religion, birth or living place, 'health/wellness', and/or their ability to satisfactorily re/produce, 'excel' and 'behave.' You do not have to be disabled to experience ableism" [77].

creatively subvert access barriers. Our paper uses misfitting to examine assistive and access technologies. Recently, Williams drew upon the misfitting framework to critique robotic interventions that are built to manage the "misfit" of autistic people, arguing for transformative futures that allow for solidarity [129]. In the context of AI VAT, we question who misfits and fits in AI systems, noting how there is a stereotypical representation of a blind person that simultaneously fits and misfits in AI VAT.

2.2 Visual Assistance Technologies (VAT)

Broadly, visual assistance technologies (VAT) are camera-based mobile applications that add to blind people's existing ecosystem of assistive tools (e.g., adding braille stickers and engaging their mobility skills [106, 113]). VAT is generally split into two categories based on how visual information is provided: either from human assistance or AI systems. Human-enabled VAT applications can be mediated by volunteers (e.g., Be My Eyes[2]), trained agents (e.g., Aira[3]), or crowdworkers (e.g., VizWiz [21]). AI VAT, such as Microsoft Seeing AI[4], TapTapSee[5], Envision AI[6], and KNFB Reader[7] [1], typically use techniques such as Optical Character Recognition (OCR) to recognize text in scanned documents, object recognition to identify products, scene description, color recognition, or facial recognition. Most recently, Be My Eyes (a human-enabled VAT) introduced Be My AI, a feature that incorporates the large language model (LLM) GPT-4 developed by OpenAI, enabling blind people to submit images, receive visual descriptions, and ask follow-up questions [96]. Prior work has studied the appropriate use cases for human-enabled VAT and AI VAT [23, 53, 54, 71]. Overall, they found that human-enabled VAT is suited for 'subjective' tasks (e.g., receiving fashion advice [28]) and complex tasks (e.g., obtaining the license plate of a ride share before entering the car [26]). In contrast, AI VAT is typically used for 'objective' tasks such as reading [92]. Although human-enabled VAT is currently more powerful, connecting to a human assistant can be time-consuming or costly [12, 41]. Thus, AI VAT remains an important tool for bolstering visual access.

However, errors are common when using AI VAT [60, 71]. Partly, this may be because foundational AI models are built without direct engagement with blind communities. In their recent review of smart assistive technologies for visual access, Gamage et al. found that 82% of studies did not involve blind people and that there is a disconnect between tasks that blind people wanted AI support with (e.g., completing paper forms) and the tasks researchers prioritized [44]. While traditional and emerging computer vision systems are typically celebrated for high accuracy rates, their accuracy rates plummet when used by blind people [52, 54, 85]. Furthermore, AI VAT may be biased since common object recognition datasets are primarily trained on objects consumed and produced in the West [35, 94, 108], and OCR performs worse when processing non-English documents [6, 51]. For instance, the accuracy of commercial OCR in Arabic is estimated to be less than 75% for printed text [5], whereas it is 99% accurate for printed English text [32].

Responding to calls to include disabled people as experts and co-designers of AI technologies, scholars have begun taking disability-centered approaches [7, 60, 69, 124] to designing AI VAT. For example, Theodorou et al. engaged blind communities in the data collection process [124], which led to increased representation of blind people in the datasets used to train AI systems. We add to this legacy by conducting an in-depth qualitative study with blind people to unpack errors they have experienced in AI VAT (e.g., processing limitations and cross-cultural bias) and highlight their perspectives for improving future AI VAT.

2.3 Understanding & Verifying AI Results

AI systems are highly opaque and stochastic, leading to potential misalignment between users' expectations and systems' functionality [37, 135]. Users may have difficulty understanding AI outputs [38] which can cause misuse [103]. To bridge these gaps, HCI research has begun unpacking how users interpret and understand AI in various domains such as social media algorithms [37, 116], healthcare [29] and creativity [31]. One promising area that aims to support user understanding of outputs is explainable AI (XAI) [78, 88], a key facet of responsible AI efforts [10, 11, 76]. However, there is no one definition of XAI. It broadly involves system transparency [48, 78] and post hoc explanations of specific model outputs by displaying heatmaps [72, 83, 114], communicating uncertainty [136], providing factual reasoning for why a specific prediction was generated [99, 100, 120] and explaining counterfactual cases [70, 88]. Previous empirical studies have demonstrated that users negotiate when and how to apply AI explanations in their decision-making process [75, 101], and their perceptions of AI explanations often depend on their technical backgrounds [36]. In addition to XAI, researchers have also called to empower users by enabling them to contest and challenge AI results [79].

Much of the work on AI explainability has focused on non-disabled people. Despite disabled people being early adopters of AI [20], we know considerably less about how disabled people make sense of AI uncertainty, and what might explainability add to (or distract from) their existing process. This is crucial considering disabled people may not always be able to identify AI errors [7, 49], leading to overtrust in AI systems [81]. A recent thread of accessibility work started to investigate how disabled people detect and negotiate AI errors. Huang et al. demonstrated how Deaf and hard of hearing (DHH) people detect false positives in sound recognition systems by validating with trusted hearing people [65]. They argued that soliciting and incorporating user feedback in sound recognition systems empowered DHH people, providing a greater sense of agency. In the context of visual access for blind communities, Abdolrahmani et al. studied blind people's reception to AI errors when navigating indoors [2]. They found that blind people's acceptance or rejection of AI error is highly contextual: errors that carried low risks were acceptable, whereas errors that may lead to socially stigmatized consequences were not (e.g., misidentification of bathroom gender signs). To mitigate these risks, Abdolrahmani et al. noted blind people would inquire sighted people and use mobility skills. Accessibility scholarship also studied error in the context of

[2]Be My Eyes: https://www.bemyeyes.com/
[3]Aira: https://visualinterpreting.com/
[4]Seeing AI: https://www.seeingai.com/
[5]TapTapSee:https://taptapseeapp.com/
[6]Envision AI:https://www.letsenvision.com/
[7]Now known as OneStep Reader: https://sensotec.be/en/product/onestep-reader/

image description and alternative text [50, 81, 134, 137, 138]. Blind people worried about potentially posting embarrassing photos on social media, and they identified various information needs that should be incorporated in automatic alterntative text such as photo quality and key visual elements to help decrease risks [137]. In formative work on the inaccuracy of alternative text, MacLeod et al. argued for designing for mistrust and communicating AI uncertainty through showcasing confidence ratings to blind people [81]. Recently, Gonzalez et al. conducted a diary study of an AI-powered scene description prototype to understand possible use cases and satisfaction with the technology [50]; they found that people generally lacked trust in its descriptions. They described that blind participants would sometimes test the AI with known images, indicating that blind people likely already have specific verification strategies that they use in practice. Building on this work, we aim to detail these tactics further, contributing an understanding of why and how blind people detect and verify errors in a variety of commercial AI VAT.

We join this relatively new line of research focused on accessible AI verification and explanations [2, 49, 81] by foregrounding blind people's everyday verification experiences of AI VAT, and their desires for XAI features that better communicate AI output.

3 Method

We take a qualitative research approach to explore how blind people verify AI VAT errors, and the limitations and possibilities of emerging explainability features. We first requested that participants share three examples of using AI VAT with us. Drawing from these example scenarios, we conducted semi-structured interviews focusing on participants' current experiences with AI VAT, how they navigate uncertainty, and their desires for upcoming features. In what follows, we detail our recruitment strategy and procedure.

3.1 Participation

We collaborated with the National Federation for the Blind (NFB) to recruit adult participants who are blind or low vision and use AI VAT. We used a recruitment survey to confirm eligibility (at least 18 years old, live in the United States, and have used AI VAT). The survey asked respondents about their preferred visual assistance technologies, age, race, and gender to ensure diversity in our sample. Over 300 participants completed the survey, and we contacted approximately 50 respondents to schedule interviews. We tried to reach out to trans/non-binary people, people of color, and older adults as their perspectives are often marginalized in HCI and accessibility research [27, 58, 122]. 26 out of the 50 respondents scheduled and completed an interview. We compensated participants with a $35 (USD) Amazon gift card. This study was approved by our Institutional Review Board (IRB).

All participants were daily or weekly users of VAT (with Seeing AI, Be My Eyes, and Aira being the most commonly used). Table 1 provides a breakdown of VAT types used by participants. To preserve the anonymity of participants, we report the demographic information of visual disability, gender, and age in an aggregated format. In terms of visual disability, one participant had low vision, while the remaining (n = 25) were totally blind with a mix of those who had light perception (n = 12) and no light perception (n = 13). Fifteen participants are visually disabled since birth, six acquired during childhood or adolescence, and five acquired during adulthood. In terms of racial/ethnic identity, ten participants are White, five are Hispanic/Latinx, four are Asian, two are Middle Eastern, one participant is African American and Native American, and one participant is Hispanic and White. Three participants did not report their racial or ethnic information. Fifteen participants identified as women and eleven identified as men. We collected participants' age range (e.g., 18-24, 25-34, etc), and the weighted average of the twenty five participants (one participant did not report) was approximately 41 years old. Participants worked or were interested in diverse fields such as technology, education, business, and art. Table 2 in the appendix describes their technical background.

3.2 Procedure

3.2.1 Pre-interview scenarios. Before conducting interviews, we asked participants to send us three examples of their AI VAT usage (e.g., Seeing AI, TapTapSee, and Be My AI). Specifically, we requested these examples to be around their confidence level (high, unsure, low) of the accuracy of AI output. Participants described recent or past interactions with AI VAT in written format or provided screenshots showing the output.

As we collected pre-interview scenarios for a couple of participants, we realized the potential of exposing participants to privacy risks and the additional work we might impose on participants to verify privacy leaks [7, 118]. We also recognized that blind people may use AI VAT for sensitive contexts despite privacy risks to address access barriers [7, 118]. We brainstormed an amendment that mitigates privacy risks while acknowledging the complexity of our request. After careful consideration, we used the following stipulation: *"In case you are planning to actively use visual assistance technologies for the sole purpose of documenting examples for this study, please do not use sensitive information such as your ID, passport, credit card, social security card, or medical information."* We hoped that this request would ensure that we did not expose participants to additional privacy harms while refraining from influencing how participants use VAT beyond this study.

3.2.2 Semi-structured interview. The first author led semi-structured interviews with blind and low vision people remotely over Zoom. Interviews were conducted from September to October 2023 and took approximately 45-60 minutes to complete. The interviews focused on various topics, including the promise and limits of AI VAT, why blind people may sometimes prefer working with sighted people instead of (or in addition to) AI, how blind people make sense of AI outputs, and what participants desired for future AI VAT. We updated the interview questions based on the pre-interview scenarios that participants shared. For example, we introduced follow-up questions on which applications participants used for different scenarios, why they thought the AI produced a particular output, and what a more appropriate output would ideally look or feel like. Interview questions can be found in the appendix. Each interview was audio-recorded and transcribed by the second author.

3.2.3 Data analysis. We followed a reflexive thematic approach [24, 25] to analyze data. After one month of closely reading transcripts

Type	Name	Description	Count	Participant IDs
Human assistance	Be My Eyes	Mobile, volunteer-based human assistance	22	P1, P2, P3, P5, P6, P7, P9, P10, P11, P12, P14, P15, P16, P17, P19, P20, P21, P22, P23, P24
	Aira	Mobile and desktop paid human assistance	16	P1, P3, P6, P7, P8, P9, P10, P12, P15, P16, P17, P18, P19, P23, P24, P25
LLM-based AI	Be My AI	Image descriptions from OpenAI's GPT-4, with chat for follow-up questions	7	P1, P4, P7, P8, P19, P20, P23
Traditional AI	Seeing AI	Mobile computer vision for reading text, recognizing color and light, and describing scenes	26	All participants
	TapTapSee	Mobile computer vision for object detection	13	P1, P2, P3, P4, P5, P6, P10, P11, P15, P16, P17, P19, P26
	Envision AI	Mobile computer vision for reading text and recognizing objects	9	P1, P6, P7, P10, P11, P12, P14, P15, P24
	KNFB Reader	Mobile OCR with text-to-speech, text-to-Braille, and text highlighting	10	P4, P6, P7, P12, P14, P15, P18, P20, P21, P24

Table 1: Summary and description of visual assistance technologies (VAT) used by our participants at the time of the study.

and open coding, early patterns included categorizing errors based on AI techniques (e.g., OCR or object detection) and surfacing blind people's error detection strategies. The first two authors refined these patterns by continuously (re)reading transcripts, writing memos, affinity diagramming, and having weekly discussions with other co-authors. We arrived at themes related to common errors blind people navigate when using AI VAT, strategies blind people use to make sense of these errors, and how they imagine future technologies to support their process of detecting and verifying AI errors. To further refine our early themes, we repeatedly returned to transcripts and searched for quotes that support, complicate, or extend these findings. In all parts of our research process, but perhaps especially during data analysis, we were aware of our positionality as sighted accessibility researchers. We acknowledge this as a limitation given our study's focus on the blind and low vision experience.

4 Findings

Our findings are organized into three sections around blind people's experiences with AI VAT outputs. First, we identify errors blind people encounter when using AI VAT. Then, we describe how blind people cultivate an intuition for assessing AI accuracy through 1) everyday experimentation with AI VAT in low-risk and known settings, 2) employing non-visual sensemaking skills, 3) collaborating with sighted bystanders and community members, and 4) cross-referencing various technologies and applications. Lastly, to support their verification and error detection strategies, participants highlighted opportunities for emerging explainability and contestability features within AI VAT. Taken together, findings unpack the technological, social, and cultural facets of navigating visual access.

4.1 Errors in Visual Assistance Technologies

In this section, participants identified common errors when using AI VAT: formatting and processing errors, as well as cross-cultural biases. We start by explaining blind people's encounters with errors when accessing complex real-world textual information, such as tables and blank lines designated for signatures. Next, we highlight cases of cross-cultural bias, noting how AI VAT fails to account for non-English languages and diverse cultural artifacts.

4.1.1 Processing Errors. Participants described several errors they encountered in AI VAT such as color misrecongition and LLM confabulation. We will elaborate on these errors in later parts of our paper. Primarily, participants reflected on processing errors where AI VAT takes raw model output (e.g., labels/detections and bounding boxes) and processes that information to make it readable for users. Given limitations with the underlying models, as well as with how this data is processed, users sometimes experience errors. For example, when using AI VAT to read documents, participants discussed uncertainties around how AI processes layouts such as date formats, tables, and blank spaces. P2, who sometimes uses Seeing AI to read work documents, described how the text '10/01' was processed as *"'You need to complete this paperwork by October 1st,' is what Seeing AI would say. When in actuality, the date was January 10th."* AI VAT encoded the date and communicated an incorrect date format (i.e., instead of following the month/day format, it described day/month). Without knowing the raw text that was recognized, this result may seem reasonable. The ambiguity around AI processing is amplified when trying to read tables. P4 explained *"take the example of this class schedule, right? So you take a picture with the document feature in the Seeing AI app. Somehow it does not recognize or it messes up because it's in a table format [...] you have to keep wondering, okay, this class and this class with so and so professor in so and so class."* These errors may occur because AI VAT does not process table contents in a manner that is accessible and usable for blind people. In a similar frustration with how Seeing AI processes table content, P12 tried following a recipe on the back of a brownie box container. However, Seeing AI's document feature would read *"a little bit of the nutrition information and a couple of lines of the recipe and then some more nutrition stuff and then some more of the*

recipe" (P12). This fragmented reading was likely because of how the text is processed. Detected characters are put into an order prioritizing text on the same horizontal line, which will logically order a single-column document, but fail for many of the real-world tasks that blind participants wanted to complete. Additionally, Seeing AI does not indicate non-text visual information on paper documents, such as blank input spaces. Some participants discussed this erasure as a form of error since AI VAT did not accurately convey important details of the document. P2 explained *"if there's a checkbox, it'll read the information for the checkbox, it won't tell me that there's a checkbox. If there's a question, it'll read the information of the question. And then move on to the next question without letting me know there's a blank there for information input."* Our findings emphasize that it is important to consider what textual, visual, and layout characteristics (e.g., blank space, table format) are valuable to blind people, and how to accessibly convey such information.

4.1.2 Cross-Cultural Bias. Some participants described OCR performance on non-English documents as a *"nightmare"* (P14). When trying to use Seeing AI and Envision AI on Korean text, P14 told us *"I couldn't read for nothing."* AI VAT also failed to support some languages in the Global South as P10 explained *"my native language, it is hard to come by. The Khmer language, the Cambodian language, it's not one of those languages that a lot of them [AI VAT] focus on."* The lack of multilingual support could also impede accessing English content. P2 noted OCR did not work when trying to understand an English-Arabic translation of the Quran, the Holy Book for Muslims. They said *"[Seeing AI] didn't know what to do with the Arabic. I think it tried to interpret it as random letters or even a picture of some sort. And then the English, because it was side-by-side, I think the Arabic confused it to the point where it didn't capture the English."* While Seeing AI unfortunately has yet to support Arabic, P2 was surprised to learn that the application's OCR would attempt to decipher Arabic text as English. Beyond language and text-related errors, some participants mentioned that Seeing AI's Product does not recognize products that are less common in the U.S. P19 told us:

> *"I have a Mexican store right next to me. It's a little store. A mom-and-pop shop that imports a lot of stuff and they have a lot of unique delicacies that you can't really find at Walmart. Ideally, I want to walk into this store and scan my way through, finding the snacks or the products that I need. But unfortunately, there have been situations where I buy some kind [of] soda or a product. I scan it and it takes me like six minutes to try to find the darn barcode and then when I find the barcode and [Seeing AI] takes its time to scan it, and it says: 'Sorry, Seeing AI cannot recognize this, please try again.' And I'm like, what the heck? I'm just going to throw my phone out the window [...] So that's just incredibly frustrating and inequitable."*

Cross-cultural errors also occur when AI VAT described images and scenery. One participant used Seeing AI's scene feature to identify their graphic tee, and it resulted in a culturally offensive misrecognition where cultural dress was instead labeled as an animal. P10 explained *"[t]he picture is supposed to be a woman. It's from my culture, [related to a] dance. It's called an Apsara. Instead of describing that the lady had a crown on, [Seeing AI described] it as an animal and [in my mind] I said: 'No, I don't have a picture of an animal on that shirt.' I have a couple of graphic tees, but I know I don't have a picture of an animal on any of my shirts."* Overall, findings affirm prior work that demonstrated how AI technologies, in general, tend to have a Western bias [6, 35, 51, 94, 108], and that AI-enabled assistive technologies often neglect to recognize the needs of blind communities from various racial and ethnic backgrounds [7, 16].

4.2 Process of Verifying AI Results

In this section, we detail how blind people verified AI VAT output. We start by noting that verification skills are developed through routine use. Then, we delve into more concrete strategies such as sensing objects as a means of verification, testing with sighted people, and switching between different devices and applications.

4.2.1 Verification Through Everyday Experimentation. Blind people described verification as an orientation they built through their experience of using AI VAT. While AI VAT is commonly advertised as 'seeing' for blind people [17, 104], P3 rejected this premise and affirmed that VAT is often wrong. P3 explained:

> *"If you think that seeing is this objective thing as opposed to interpreting, then that's a flaw in itself, and it is imbued in these [VAT] apps, right? Like, oh, we're seeing this for you because you can't, and so we're going to tell you what it is. And I think as a blind person, at least for me and some of the folks that I know, we know some of this is wrong."*

Challenging the misconception of equating AI VAT to normative sight, using these applications requires building familiarity and constant negotiation when interpreting results. As such, when we asked participants what advice they would give to blind people who are just starting to use AI VAT, most participants emphasized exploring these applications, noting that *"[AI VAT] are fun. Just play with it. Don't rely on them. Find their strengths and weaknesses"* (P20). P24 added *"make sure that you're not testing it on something that is going to impact your life in a negative way if the information you get is incorrect."*

Through experimenting with AI VAT, participants discussed how AI VAT is often used for non-visual tasks where they already have a sense of what the correct information should be or they could easily verify. Reflecting on their experience, P7 said *"I'm still very much kind of testing for myself what I can rely on. So, I've been very careful. I mostly use [AI VAT] in low-risk situations where either I already have the data [...] I've been a little bit more reluctant to completely rely on it for my first and only source of information."* For example, P7 shared with us an incident where they dismissed AI VAT output. They used Be My AI in a bar they frequent and know very well. However, Be My AI produced incorrect results. P7 said *"[Be My AI described brands] I knew for a fact that they didn't sell."* In trying to reason why this error occurred, P7 illustrated the logic of Be My AI as *"if there's a tequila, there must be another tequila next to it. And I can't tell which tequila it is, so let me just make one up."* This approach of exploratory use (i.e., testing AI VAT in known contexts) constructed a perception of AI VAT as *"more of a guidance thing than a full support thing"* (P26). AI VAT becomes a mean to

get surface-level information rather than details, and this critical use is shaped by prior experience.

Participants avoided using AI VAT for visual tasks that are difficult to verify, and have a high probability of error. For example, *"tasks such as maybe finding something that's been dropped or something more complex such as descriptions of pictures or looking at clothing to see if there are any stains"* (P1). While participants described some tasks (e.g., reading a document) as frequently containing errors, they still used AI VAT since such errors are easily detected and resolved by context. These errors were referred to as *"classic OCR mistakes"* (P11) which involved character misrecognition. P12 explained that AI VAT would scan the letter `m` as the letters `rn`, noting *"[OCR] does that often with web addresses. It will say 'dot corn,' and I know it's 'dot com.'"* Similarly, P25 added *"I've never heard of Heetos; I'm pretty sure that meant Cheetos."* Because of the frequency of OCR errors, some participants mentioned that they have grown accustomed to *"reading between the lines"* (P16) when using OCR. P12 explained *"I'm so used to accounting for it without even thinking about it. It's like second nature to me."* In essence, participants do not always need to verify every AI result, especially in cases where they have built familiarity and can correct errors on their own. Overall, in the case of AI VAT, blind people were critically aware of limitations, and that in turn shaped their skepticism and orientation toward verification based on use cases.

4.2.2 Sensory Verification. Some participants employed non-visual senses (such as feeling and hearing) as a tool to verify AI outcomes. P16 said *"I can shake a can of beans and a can of corn and can of green beans, or tomatoes, and I can kind of tell by the sound they make."* Some participants described their tactile recognition technique as *"usually when I touch something, I immediately know what it is"* (P13). Tactile recall depends on prior experience with the object of interest. P19 said *"muscle memory of packages or bottles that I know that I've used."* For example, P11 told us that one time an AI VAT described an item as *"possibly ice pop or possibly ice cream bar."* However, P11 said it was *"a 3 pack of the corn on the cob that you get in the supermarket. Now I knew that was corn on the cob by feeling the outside of the package. It had cellophane around it just like a pack of meat."* In the context of verifying a clothing's color, P15 told us *"I was trying to find a particular dress shirt. [Seeing AI] was saying that it was light blue, but I was pretty sure by the material that it was the lavender one."* By touching the material of their shirts and finding distinctive patterns (e.g., logos), some participants were able to refute AI's color recognition.

4.2.3 Verifying & Testing With Sighted People. Blind people sometimes involved sighted people to *"figure out where the limits of AI and what can [they] feel pretty confident about"* (P25). For example, P25 verified some AI output with their sighted partner. Through their interaction, P25 was able to understand what types of content work well with OCR and what does not, concluding that OCR does not work well with *"more colorful [packaging]"* (P25). Some participants explained making plans to reach out to trusted sighted people (e.g., friends or family members) in high-risk situations where accuracy and security are needed. P23 said *"if I'm really kind of twinging that something's not right and it's something important, I will go seek out sighted assistance."* Sighted people may also play a role in mitigating security leaks by helping with precise camera aiming before using AI VAT. P4 explained *"I have to check how much I have to pay for my utility bill. So, in that case, I have asked my dad where exactly should I point my camera because it's always gonna be the same on that page without it revealing like my account number and everything."* In low-risk settings, participants also described sometimes validating with sighted people when convenient. For example, P19 used Be My AI to get a sense of their staff room, and it described a vending machine. This felt weird to P19 since they *"worked in different schools, in different settings, and have not encountered vending machines in the staff lounge."* P19 confirmed when *"somebody just opened the door and they're having conversations with other teachers and when they're done I asked 'Hey, is there a vending machine?"'* Through this quick and mundane interaction, P19 verified that there was indeed a vending machine. Similarly, P6 verified AI VAT's color recognition with sighted people as they proceeded with other activities (e.g., en route to university). They said *"if I'm getting picked up by paratransit, it is a service that takes people with disabilities to certain locations, I'll be like, 'hey, what color is my shirt? TapTapSee or Seeing AI said this one color, I'm still a little skeptical.' [...] I'll periodically ask just to validate my inquiry."* Corroborating prior work, we found that blind people sometimes verify results with sighted people [138]. We described how our participants confirmed AI outcomes with sighted people as a way to gain a better understanding of the limitations of AI. They also deliberately engaged with sighted people for high-risk tasks that require accuracy and security, whereas low-risk scenarios were verified spontaneously with sighted people as-needed.

4.2.4 Cross-Referencing With Different Devices & Applications. Blind people discussed switching between different devices and applications to verify AI results. In particular, they verified AI outputs by trying to confirm consistency. For example, some participants told us they would cross reference with *"old school"* (P17) technologies like portable scanners. P4 said *"I get a lot of letters from my college, so I skim those [using Seeing AI] [...] I also have a portable scanner. I do not know the name of the scanner, I scan through that also."* One participant explicitly hoped that VAT applications would support blind people's reflection and verification process. P8 wanted a way to save and *"consolidate the information in the chat feature [of Be My AI]"* so that they *"could go back and compare [and] have that as a reference point."* Participants also switched between different features within one application. P2 tried to read a paper using Seeing AI's Document mode. However, it produced an error message of *"No image visible,"* (P2). To verify this issue, P2 *"switched to Short Text [another feature in Seeing AI], then it recognizes text."* Nevertheless, it is difficult to read long forms of text using Seeing AI's Short Text mode since users have to keep the camera stable otherwise it would *"just start reading from the top again"* (P21).

While switching between different devices or applications gave blind people an avenue to validate AI, it did not always lead to a resolution. P26 took a photo of a room to get a sense of what was around them. TapTapSee and Seeing AI produced conflicting responses. They explained *"the uncertainty came from the different wording in the descriptions like clock versus hat. Is it a clock? Is it a hat? Is it even really one of those things? [...] the specifics is where it got a little bit different. It didn't always overlap and so that's where the uncertainty came from."* Similarly, P1 was *"looking for a*

particular setting" in their grill. P1 already had a sense of what the output should be, but they wanted to double-check using Be My AI. However, P1 said *"when [Be My AI] gave me the order [of buttons], I could tell from that description that was not correct at all. There were buttons that were out of order."* P1 could not cross-check with Seeing AI *"because those are symbols that are on those buttons, not actual texts. Seeing AI would not have read those very well. It's designed to read actual text."*

Furthermore, participants may sometimes cross-reference in creative ways, using devices that are not assistive technologies. P19 used Be My AI to read their work phone's extension number so they could *"give my family this number in case of emergency."* After receiving the extension number from Be My AI, P19 told us *"I called this number because I just wanted to verify. If my office phone rings, great. If it doesn't, I don't know what's going on.[...] I got straight through the school counselor's phone line and I was like 'okay, no, this is not it.'"* P26 added that when they use Seeing AI Product mode and it outputs an unfamiliar brand, they search for the brand online. They explained *"if it gives me a brand that I've never heard of and I don't remember buying this. I can always type it in online and verify that it is in fact a brand for this product"* (P26).

4.3 Blind People's Perspectives on AI Explainability & Contestability

Participants explained that it's inaccessible to non-visually interpret the images captured and processed by AI VAT, broadening XAI efforts beyond AI outputs. Then, participants discussed how XAI approaches, such as conveying confidence ratings in AI VAT outputs, may hinder and support their verification process. Lastly, participants articulated that AI VAT should incorporate avenues to challenge and contest AI VAT output.

4.3.1 Interpreting Input by Interactive Camera Guidance. Participants discussed how errors and uncertainties may emerge because it is inaccessible to know what was inputted into AI VAT. As P23 explained, *"sometimes I wonder whether or not the inaccuracy is really more my picture taking skills than actual inaccuracy."* Blind people may not know what was captured by their camera [3, 33, 66] and processed by AI. Participants thought that some errors could be *"user error, [that is] not pointing the camera properly"* (P16). For example, participants said their images could be blurry or too close to the object, resulting in faulty AI VAT output. Past work also demonstrated how the lack of camera guidance techniques can be inaccessible to blind people, and could lead to inaccurate AI output [63, 138]. Additionally, some images *"can be very difficult to read. [...] if you have something neon, and it's got white text that could be very difficult to read for anybody whether sighted person or computer"* (P18). Accordingly, participants desired affordances that enabled them to understand image quality as it impacts AI performance. Some Be My AI users mentioned that this type of feedback is sometimes given. P4 said *"[Be My AI] also tells you if I didn't take like the picture properly, or if there is an obstacle in that picture."* However, participants hoped that Be My AI would provide feedback in real-time (as a user is taking the photo), and suggest how to better retake the photo. P1 explained *"if there was a way to tell the user how to improve the picture in future."*

Some AI VAT such as Seeing AI and KNFB Reader (now known as OneStep Reader) have built-in support to help guide blind people in capturing specific areas of interest. However, these current features are insufficient. P23 described existing guidance techniques as in their *"infancy stages."* P12 added *"the [guidance] technology is there, but it's far from perfect."* Overall, current camera guidance features are thought to be *"sometimes helpful, sometimes it's kind of a pain in the butt. Not because it's wrong... They're just very visual"* (P25). For instance, participants found Seeing AI's guidance in Document mode particularly frustrating since it only indicates the document edges, leaving blind people to guess which direction they should move to correct the view. P10 explained *"[Seeing AI] doesn't say move to the left or right. You kind of have to guess. If it says top left corner, that means your camera's kind of leaning more towards the left, so you kind of move it a little back to the right."* Besides the perpetual need to remember to move in the opposite direction (i.e., going right when it indicates left edge detected), Seeing AI's guidance can be inaccessible because it fails to specify how much users need to move. P25 explained, *"it feels like a game. [...] I'm thinking there are all these degrees. It's not just upright left. It is not just super simple like that. To what degree do I go down? So, when [Seeing AI] will say 'left edge visible.' I'll go down and it'll say 'left edge not visible' and I'll go down and It'll say 'right edge not visible' and I'm like, [imitates screaming] you know."* Seeing AI's guidance is further difficult on a round object. P10 tried to use OCR on a prescription bottle and said *"it is hard just because of the way the bottle's made because it's round so you have to move the bottle for it to read what it's saying."*

In imagining how future VAT could improve, participants discussed opportunities to further refine camera position guidance. P14 reflected on their experience with Aira, a human-enabled VAT, and how they wished the guidance of AI technologies was similarly *"interactive."* They said *"I go to Aira, I want to do like a team viewer session with Aira. So I pointed my iPhone camera at my computer screen and I asked [the Aira Agent], 'can you see the IDM password on the team viewer?' They tell me 'oh, go more to the left, go more to the right. Closer to your screen.' That's a lot more helpful than just 'Oh, right edge detected, left edge detected [Seeing AI's guidance]' [...] So yeah, I wish it was more interactive like that."* While redesigning an entire guidance system that is akin to interacting with sighted agents may be ambitious, a simple change to Seeing AI's guidance could be *"just focusing on where you want the camera to be instead of where you don't want it to be"* (P8), eliminating an added cognitive burden. Participants also wanted the guidance to be based on shape. For example, when trying to perform OCR on cylindrical objects (e.g., bottles), VAT *"could say something like 'rotate right' or something like that as it's reading. Because maybe it would sense that the words were cut off and then the object was like a cylinder"* (P21).

Overall, participants discussed how it might be difficult to verify AI VAT output because it is inaccessible to know what was captured by their cameras and processed by AI. While some AI VAT does incorporate camera guidance features, participants felt these existing cues were insufficient. They envisioned future camera guidance to more interactive and less cognitively burdensome.

4.3.2 Tensions Around Communicating Accuracy. Participants wanted to understand why AI produced a specific response. While prior

work proposed incorporating confidence ratings in automatic alternative text [81], participants in our study had mixed opinions about the benefits of indicating uncertainty in AI VAT. Some participants felt it was important to convey a level of uncertainty through quantifiable measures or phrasings of output. For example, in the context of outlandish misrecognition, P9 used an AI VAT to describe their cat and the application *"said 'rattlesnake.' I was like, okay... That's pretty hilarious because obviously there's not a rattlesnake here!"* When asked whether it would be helpful to get a sense of why AI generated this description. They elaborated *"so if [AI] said like. 'Yeah, 10% rattlesnake.' It would at least give me some sort of nuance to the fact that [...] I know this isn't a rattlesnake. But there's something about the color of her fur that maybe looks like a rattlesnake."* In the context of using AI VAT to read documents, participants described how getting a sense of accuracy would save them time. P20 told us *"The app can tell you, 'well, this is like 80% accurate or 90% accurate' Then you will know, it's not it's not a good OCR, so you might need to go find someone to read it to you."* Preferring less quantifiable indicators, P2 explored verbal descriptors and emphasized the importance of indicating where in the document that uncertainty lies. They said *"I don't know what phrase could be used, 'Uncertain,' 'Less certain,' or 'Possible.' Something like that. But especially indicating where that text is on the document or at least having a marker for it."*

Overall, some participants argued that it is critical to convey uncertainty, especially in LLM-enabled VAT (e.g., Be My AI) which often has humanistic undertones that might encourage overtrusting the output. P23 explained, *"I worry about the misinformation problem. In the sense that If it suddenly sounds human and if it's giving you all of this reputable sounding information. Are you going to take it more realistically because it sounds human rather than a robot?"* However, one participant thought that communicating accuracy is an unreliable measure since AI VAT does not have ground truth. P22 elaborated:

> *"I don't see how that would be feasible or practical [...] because [AI] is relying on its own measurements. So unless it has some way to measure against the original document and compare the outcome of the picture and OCR it performed, there's really no way to calculate those statistics."*

Few participants did not have strong opinions about communicating uncertainty in AI VAT. Some of these participants felt that they take a utilitarian and task-based approach to AI. When they experience an error in AI VAT, they do not want to waste time trying to understand the issue. P25 explained *"I guess it doesn't seem relevant for me. Either it's working or it's not working. I don't really need too much information about it. I got too much stuff going on."* One participant noted how these additional explanations can be problematic, especially during moments of AI bias where other actions are more appropriate. P23 said:

> *"There are a lot of experiments that have been done around like AI and determining individuals of different races and different ethnicities and how it pretty consistently gets certain things wrong. Using that as an example, I don't really care why [AI] got it wrong. I cared that it got it wrong. Fix it. Make it better. Make it right."*

4.3.3 Contesting AI Output. Participants felt that the existing AI VAT systems do not reflect their preferred way of negotiating visual access. P3 told us that AI systems do not allow for shared *"dialogue or discourse."* Reflecting on their process of building visual access with sighted people, P3 said:

> *"Well, I think when you have a [sighted] person, whether it's via FaceTime or Aira or in person, you have a back and forth [conversation] between that person [...] So if somebody reads to me something and I have a form to complete, I'm not going to just let that person take over. It's going to be back and forth."*

P23 added *"with a human not only are they going to give me a description, but I can interrogate that description, right? Okay, so you say it's blue. Well, is it dark blue? Is it light blue? I can refine."* These quotes emphasize that blind people play a critical role in shaping access [15] as demonstrated by participants' active negotiation with sighted people. Some AI VAT limit blind people's ability to contest its output (i.e., challenge AI systems [9, 79]). As P20 explained, *"you cannot really ask [Seeing AI] a question because you take your picture and you get the description and that's it. So, you cannot really interact with [Seeing AI]."* However, participants shared strategies they currently enact to contest and improve AI VAT output. They also articulated potential avenues for future VAT technologies.

Some participants used Be My AI's chat feature to contest and teach AI. Recalling the incident where Be My AI misrecognized their work phone's extension number (shared in 4.2.4), P19 said *"I can also type in my feedback, say 'Hey, this is not the extension number.' It comes back saying 'sorry, I apologize for the mistake.' So I always like to do that because then as a user I have the power to improve it."* When asked why they think it is important to provide feedback to Be My AI during chat sessions, P19 elaborated *"I really like giving feedback to the AI-powered models more often because they're AIs, they're computers and they sometimes hallucinate and make mistakes. That's why I really like to give feedback to those types of apps than others because they have potential. They are our future and we can't deny it."* There is a sense that by fine-tuning AI during chat sessions, AI systems would eventually "learn" and become better. However, that is not always the case. P20 used Be My AI to get a visual description of their photo on a nature trail. Be My AI incorrectly described their white cane as a *"hiking pole"* (P20). They followed up by asking *"is it a hiking pole or a blind cane or a white cane?"* Be My AI continued to describe their white cane as a hiking pole. P20 explained *"normally with the online platform you can tell ChatGPT the answer is not right. This is the correct answer. [ChatGPT] is not gonna argue with you, right? So it just says 'okay, yeah' So I'm not sure yet how I can do that with [Be My AI]."* In this case, the AI VAT did not immediately improve its output even after an attempt at contesting.

Participants wanted accessible and usable interfaces for blind people to contest output and provide feedback. In imagining how to incorporate feedback opportunities in AI VAT, P21 said *"just a little pop-up and it says, 'how did we do?' and [users] would have the option to skip it if [they] didn't have time for that."* One participant, P23, emphasized that feedback requests ought to be accessible and usable

by blind people who are multiply disabled. P23 explained that VAT companies should be mindful that *"blindness is not a single disability population [...] cross-disability perspectives are not a common thing in [...] typical blindness apps."* Indeed, erasing the experience of multiply disabled people is unfortunately a common occurrence in assistive technologies [80]. When designing feedback forms, P23 elaborated that developers need to consider questions like *"what methods do you have for folks who don't type but who use like augmentative communication aids?"*

Some participants voiced concerns about providing feedback to AI systems. Primarily, participants worried about being constantly bombarded by feedback requests. Reflecting on Be My Eyes feedback prompt (thumbs up or down icons) after interacting with sighted volunteers, P12 said *"when I end the call and I hit thumbs up, it's not a big deal, but I use Seeing AI so often that if every time I finished a scan or something, and it [requested feedback], it would drive me crazy."* P12 explained that it is important to have frequent feedback prompts when interacting with sighted volunteers because it may keep volunteers accountable whereas it is unclear how to *"to hold an automated app accountable."* In part, this could be because some errors are indicative of complex issues such as *"there's a bug in the app"* (P12), and it would be frustrating to repeatedly point out this problem. Instead, participants opted for less frequent feedback engagements. P9 felt that it is more meaningful to give direct feedback by conducting compensated interviews or surveys because writing feedback is work, and it should be valued as such. They asserted:

> *"It's a lot of mental energy to have to go into the app and find the 'give us feedback' thing. Then, I gotta get out my Bluetooth keyboard because it's gonna take forever to type something on my screen or I know that the voice dictation is gonna mess something up [...] pay me well to sit down and honestly give my time and energy, and I will."*

5 Discussion

Our results reveal the errors blind people encounter with AI VAT, the strategies they use to confirm or reject AI outputs, and opportunities for supporting accessible XAI. We turn to the feminist disability framework of misfitting [47], a concept produced by disability studies scholar Rosemarie Garland-Thomson to rethink disability and access. Situated in feminist literature on vulnerability and dependence, Garland-Thomson employs the metaphors of "misfitting" and "fitting" to demonstrate how embodied experiences are constructed by material relationships between bodies and worlds, seemingly comfortable when there is a fit and discordant when there is a misfit. Misfitting and fitting arise because of dynamic interactions between disabled people and structures. However, misfitting is *not* an inherent quality of disabled body-minds. For example, Garland-Thomson recounts the experience of wheelchair users trying to access a building: they may fit when there are elevators, and misfit when there are only stairs. Misfitting, with the injustice and violence it often accompanies, spotlights disabled creativity and ingenuity. For Garland-Thomson, misfitting becomes a site to affirm disabled ways of knowing.

We build from this theory to illuminate how misfitting unfolds in access technologies. In the context of AI VAT, misfitting refers to errors that arise either from technical structures or mismatches with how blind people use these systems. The concept of misfitting is an analytical lens that interrogates what it means to fit or misfit into a system. Misfitting builds on long-standing traditions in accessibility scholarship to critically examine assistive technologies. The overall goal is not to make marginalized people "fit" into technologies. Historically, that often encouraged predatory inclusion practices [17, 123]. Rather, the misfitting framework offers language that demonstrates how blind people creatively adjust their behaviors to verify inaccessible AI, while also calling into question the concept of "fit" and whether they should have to or want to do that verification labor.

5.1 Who 'Fits' & 'Misfits' in AI Systems?

The misfitting framework enables us to interrogate assumptions about fitting and misfitting in our world [47]. We use the concept to argue that blind people experience misfitting in AI VAT as computer vision prioritizes sighted camera aiming. Blind people with marginalized identities and want to access non-English text and non-Western materials experience further misfitting.

5.1.1 Misfitting in Computer Vision Systems, and the Inadequacy of Camera Guidance Techniques: In the context of AI, foundational computer vision systems are often trained on sighted people's data, making them optimized for sighted camera aiming. When used by blind people, the accuracy of computer vision systems decreases [54, 85]. In other words, there is often a *fit* when sighted people use computer vision, and a *misfit* when used by blind people. Consequently, AI VAT embedded camera guidance features in hopes of making blind ways of camera aiming more legible to AI VAT (i.e., addressing the misfit). However, we learned from our participants that these guidance features are far from adequate. Current camera guidance systems not only maintain the existing misfit but also contribute an additional layer by potentially centering sighted logic. As noted in section 4.3.1, the current guidance is not aligned with how blind people prefer to receive directional information (e.g., P25 reflected on how language is visually-centric, and P10 articulated the additional cognitive labor to make sense of where to orientate their camera based on cues). These communication breakdowns between participants and camera guidance cues could be because such features do not incorporate what blind sociologist Siegfried Saerberg coined as "blind style of perception," denoting sensory schemes of interpretation that blind people develop in relation to spatial and embodied processes [105]. Instead, current guidance cues might be based on sighted styles of perceptions, and this clash may occur because blind people cannot negotiate what type and level of guidance is more meaningful to them. A one-size-fits-all approach to camera guidance in AI VAT fails to capture the diverse ways blind people wish to receive directional camera cues based on their access needs in the moment. Our analysis offered preliminary insights into making these camera guidance cues more interactive and malleable based on object type (recall P21 and P10 notes on using AI VAT for round objects). We encourage future work to build from our study and prior work on blind photography in general [3, 33, 66] to design better camera guidance features for AI

VAT. For instance, upcoming work could explore camera guidance techniques that can detect object shape, and provide directional cues with length indicators (e.g., specific measurements or verbal descriptors) or guidance on how much to rotate round objects.

5.1.2 Marginalized Blind Communities Encounters With Misfitting: Our analysis demonstrates how AI VAT may privilege the experiences of an imagined 'typical' blind user, neglecting those who do not fit this ideal. In thinking about who is seen as a 'fit' to mainstream society, Garland-Thomson asserts that to "dominant subject positions such as male, white, or heterosexual, fitting is a comfortable and unremarkable majority experience" [47]. That is, experiences of misfitting are often racialized and gendered. Building from our findings and prior work [7, 16, 58], we argue that blind people who hold marginalized identities (e.g., being an ethnic minority in the US) misfit in general-purpose AI VAT and computer vision systems, as previously discussed. AI VAT are often marketed as 'international' and general purpose, attending to the needs of blind people around the world [104]. However, our findings complicate this narrative. In section 4.1.2, participants noted instances of cross-cultural bias, articulating how AI VAT do not attend to their cultural products, artifacts, and languages. P19 shared their frustration of being unable to use Seeing AI in a local Mexican store. P10 felt that their native language, Khmer, is not of interest to AI VATs. These cultural and linguistic erasures may lead to multiple types of harm as described in Shelby et al.'s taxonomy of algorithmic harms [110]. For example, it may result in quality of service harms [22, 110] because of performance discrepancies between blind people who speak English and prefer Western products and those who do not primarily speak English and consume non-Western foods. In the emerging research area of fair and just AI for disabled people [17, 93, 128], upcoming work must challenge the tendency to think of disability as a singular identity and recognize how intersectionality [19, 34, 58] critically shapes experiences of AI error. What we, AI and accessibility researchers, define as universal or general purpose is likely removed from reality and perhaps reflective of Western norms [61]. Nevertheless, there is an exciting trend in research that aims to empower blind people in shaping personalized object recognition [4, 68]. We hope this translates to commercial AI VAT, in addition to improving cross-cultural OCR and image description.

5.2 Misfitting Celebrates Disabled Ways of Knowing

Garland-Thomson argues that disabled people, through their experience of misfitting, cultivate expertise in building access [47]. By experiencing misfitting in a particular time and space, disabled people gain a sense of resourcefulness to circumvent inaccessibility. For example, she notes that blind people uniquely navigate the world without relying on vision, a skill sighted people lack. Specifically, blind epistemology, as Caroline Jones writes, "demands a rethinking of how we form knowledge" [67]. In turn, this shifts the tendency to frame blindness and disability as a "problem" to a source of creativity and a position of knowing.

Despite the lack of affordances to support verification in AI VAT, our findings emphasize the creative ways blind people confirmed or rejected AI VAT. Contrary to the dominant discourse around AI VAT as "seeing" for blind people [15, 104], participants challenged painting sight and AI as objective forms of truth (recall P3 quote in section 4.2.1), affirming blind people's critical role in the process. By applying the misfitting commitment of emphasizing "remediation over origin" [47], our analysis reveals how blind people use complex and ordinary methods to make sense of AI when faced with error and uncertainty. In particular, findings described how blind people engaged in intrasubjective and embodied practices. For instance, in section 4.2.2, participants employed non-visual sensemaking, like feeling or hearing objects, as a means to dismiss AI VAT. Complementing Gonzalez et.'s research on scene description [50], we learned that blind people test AI VAT in low-risk and known contexts to inform future use. Our participants also discussed their gained familiarity with reading text filled with *"classic OCR mistakes"* (P11 in section 4.2.1). Additionally, participants described intersubjective ways to verify AI, such as strategically engaging sighted people. Building from prior work on interdependence [15, 90], which emphasizes how access is co-created by both disabled and non-disabled people, our findings highlighted the mundane and deliberate ways blind people included sighted people to verify AI. Taking these strategies together, we emphasize that the verification process undertaken by blind people is not additive. Rather, it is a critical step that enables visual access.

Overall, failing to support blind people's existing verification efforts within AI VAT applications may potentially weaken genuine AI partnerships, a value that disabled communities hold as noted in our findings and prior work [43, 60, 93]. A misfitting lens attunes us, accessibility and AI researchers, to the creative epistemic labor developed by disabled people. We advocate for designing AI VAT systems that are aligned and in harmony with blind people's verification process. For instance, P8, in section 4.2.4, proposed that AI VAT should allow users to save outputs to support blind people in cross-checking with other applications. Using our findings as a starting point, we invite researchers to co-design AI VAT verification affordances with blind communities. Future work could also empirically extend this line of inquiry by developing typologies with blind people on the types of AI uncertainties they face in various contexts and how their verification strategies may evolve as new models emerge.

6 Lessons & Directions for Responsible AI

In this section, we outline specific takeaways for Responsible AI, an evolving domain in both research and industry that foregrounds principles like transparency, explainability, and inclusion [10, 11, 76, 133], within AI VAT teams and beyond. Particularly, we call on Responsible AI practice to 1) prioritize accessible XAI, and 2) work toward disability-centered audits.

6.1 Extending and Moving Beyond Explaining AI Outputs

Findings provided insights into blind people's preferences for framing AI explanations. Participants articulated how explainability may sometimes support their verification strategies (e.g., when using OCR, understanding how AI is processing complex layouts such as tables is useful). Our participants' accounts also affirm prior work on the limitations of confidence scores [7, 81]. While some found confidence scores helpful in certain contexts such as getting a sense

of the accuracy of long documents when using OCR, participants questioned the validity of these metrics since they were unsure of how accuracy is computed (as P22 explained in section 4.3.2). Indeed, Gilpin et al. argued that explainability alone is not enough; it ought to be coupled with the ability to articulate outputs, answer users questions, and be subjected to auditing [48]. Future work could study whether factors such as describing how confidence scores are calculated would help blind people negotiate trust.

Prior research on computer vision explainability may assume that users' interpretation of visual input is trivial or self-evident. Recently, Hong et al. explored the potential and limitations of providing quality descriptors (e.g., blurry image) in a teachable object recognition prototype for blind people [63]. They found that participants benefited from such information to further iterate and improve AI performance. Our study similarly demonstrates that blind people value understanding the quality of their images even in contexts beyond training object recognition (e.g., OCR). We argue that explaining AI input – especially when coupled with accessible camera guidance – serves as a site to support verification, potentially resolving uncertainty around the source of error. This empowers blind people to understand whether the error they encountered was due to poor image quality issues or model limitations.

Building from human-AI design guidelines [10, 133], our findings teach us that Responsible AI scholarship should advocate for 1) providing transparent explanations on how confidence ratings are produced (i.e., explaining the explanations), and 2) challenging ability assumptions [132]; describing key attributes of the input that may be inaccessible to users.

6.2 Supporting Disability-Centered Audits

In imagining better futures, some participants envisioned more direct engagements with AI VAT. For instance, our participants emphasize the need to contest AI outputs through feedback. However, the majority of AI VAT are closed and proprietary which is a stark departure from its participatory[8] origins. *What might it mean to include blind people's perspectives in the development of AI?* Recently, accessibility scholarship has suggested developing AI audits with disabled people to address harms [43]. Our findings offer some insights into how blind people can be involved in improving AI, and pave the way toward disability-centered AI audits. Here, we call for flipping the script of designing AI to "replace" or "extend" blind people's abilities, and towards enabling blind people to inform AI [18, 20, 125]. Some participants felt excited and even compelled to be actively involved in reshaping AI (recall P19 comment in section 4.3.3). While some AI VAT are open to receiving feedback (e.g., Be My AI noted in their blog: "[...] please be patient, and keep telling us about your experiences, positive and negative, so we can make this the best possible tool for you" [39]), it is unclear where or how to provide feedback. Participants envisioned providing feedback on the spot (e.g., in Be My AI's chat), and through formal sessions and surveys. They also emphasized equitable compensation for their work in identifying errors, and accountability measures describing how their feedback would be implemented. One way to help blind people collectively contest and repair AI VAT could be through disability-centered audits that gather their experience on AI outputs that can be improved, and objects that require further AI training. Specifically, in designing these audits, AI VAT applications may draw inspiration from ACCESS SERVER [87], a design research project that anonymizes and compensates disabled people for finding access barriers in cultural institutions. Building from a legacy of disability activism and scholarship, ACCESS SERVER affirms the agency of disabled people, makes the feedback process more accessible by providing optional templates, and values their labor by financially compensating them. AI VAT can adopt such practice within applications by *clearly* stating how their users data will be used and handled, compensating users for their data, and reporting on any changes as result of their data. Furthermore, our findings on cross-cultural bias assert the importance of reshifting our focus on quality and accuracy when auditing AI VAT, paying particular attention to cultural representation.

[8] In tracing the precursor and foundations of OCR technologies (known as optophone), scholars asserted that blind people were not merely testers of such technologies; they were also co-developers by offering recommendations of hardware and demonstrating how its used to the public [30, 89].

7 Limitations

We had several limitations related to recruitment. We tried to recruit a diverse sample of participants in terms of age. However, our current sample only included one older adult participant (i.e., over 65). Our findings may not capture the experience of blind and low vision older adults. While our sample did include participants who have diverse cultural backgrounds and speak multiple languages, our study is based in the U.S. and may not extend to the majority of the world. We also recruited participants at a critical stage in the VAT technology scene: Be My AI was only open to beta users, and we included some ($n = 7$) participants who have used Be My AI. Other VATs (e.g., Seeing AI, TapTapSee, and Aira) recently incorporated LLM features after we concluded our study, which may impact blind people's future experiences with them.

8 Conclusion

AI systems will never be 100% accurate. While there have been substantial efforts to recognize and address AI errors, most of those efforts have ignored blind people, resulting in verification processes that are inaccessible to them. This qualitative study sheds light on blind people's use of visual assistance technologies to non-visually verify AI outputs. We found that blind people often experience errors when using AI VAT to read complex layouts, and we detail cases of cross-cultural bias. To verify AI VAT results, blind people employed various tactics such as experimenting in low-risk contexts, using non-visual sensemaking skills, strategically including sighted people, and cross-referencing with other devices and applications. To enhance AI VAT, blind people desired more interactive camera guidance to negotiate AI errors. Some participants complicated common XAI techniques such as confidence ratings, noting ambiguity around how these indicators are computed. Instead, they emphasized avenues to directly contest and improve AI outputs. Extending our findings, we applied the feminist disability framework of misfitting/fitting as a generative perspective. We argued that blind people "misfit" in computer vision systems whereas sighted people fit. Blind people who are racially or ethnically marginalized experience an additional layer of misfitting since AI VAT does not account for their lived experiences. Furthermore, we called

attention to the creative ways blind people negotiate misfitting in AI VAT systems that often invisiblize their verification work. Finally, we offer provocations for the field of Responsible AI, underscoring the need to prioritize accessible XAI and work toward disability-centered audits.

Acknowledgments

This research would have not been possible without the time and expertise of our participants. Thank you to the 26 participants of this study for sharing your insights with us. We thank the National Federation of the Blind (NFB) for circulating our recruitment survey. We also appreciate Aashaka Desai, Andrea Jacoby, Cami Goray, Jessica Flores and Sam Ankenbauer for providing generous feedback and comments on earlier drafts.

This material is based upon work supported by the National Science Foundation under Grant 1763297 and by a Google Research Inclusion Award. 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 National Science Foundation.

References

[1] [n. d.]. KNFB Reader. https://nfb.org/programs-services/knfb-reader. Accessed: April 20, 2024.

[2] 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.

[3] Dustin Adams, Lourdes Morales, and Sri Kurniawan. 2013. A qualitative study to support a blind photography mobile application. In *Proceedings of the 6th International Conference on PErvasive Technologies Related to Assistive Environments*. 1–8.

[4] Dragan Ahmetovic, Daisuke Sato, Uran Oh, Tatsuya Ishihara, Kris Kitani, and Chieko Asakawa. 2020. Recog: Supporting blind people in recognizing personal objects. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–12.

[5] Salah Alghyaline. 2023. Arabic Optical Character Recognition: A Review. *CMES-Computer Modeling in Engineering & Sciences* 135, 3 (2023).

[6] Y Alginahi. 2017. Document processing and Arabic optical character recognition: A user perspective. *International Journal of Computer Science and Information Security* 15, 12 (2017), 86–97.

[7] 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.

[8] 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.

[9] Marco Almada. 2019. Human intervention in automated decision-making: Toward the construction of contestable systems. In *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law*. 2–11.

[10] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N Bennett, Kori Inkpen, et al. 2019. Guidelines for human-AI interaction. In *Proceedings of the 2019 chi conference on human factors in computing systems*. 1–13.

[11] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador García, Sergio Gil-López, Daniel Molina, Richard Benjamins, et al. 2020. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information fusion* 58 (2020), 82–115.

[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] BBC News. 2015. Google apologises for Photos app's racist blunder. *BBC News* (30 June 2015). https://www.bbc.com/news/technology-33347866 Accessed: April 18, 2024.

[14] Chris Bell. 2010. Is disability studies actually white disability studies. *The disability studies reader* 5 (2010), 402–410.

[15] 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.

[16] 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.

[17] Cynthia L Bennett and Os Keyes. 2020. What is the point of fairness? Disability, AI and the complexity of justice. *ACM SIGACCESS Accessibility and Computing* 125 (2020), 1–1.

[18] 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.

[19] Patricia Berne, Aurora Levins Morales, David Langstaff, and Sins Invalid. 2018. Ten principles of disability justice. *WSQ: Women's Studies Quarterly* 46, 1 (2018), 227–230.

[20] Jeffrey P Bigham and Patrick Carrington. 2018. Learning from the front: People with disabilities as early adopters of AI. *Proceedings of the 2018 HCIC Human-Computer Interaction Consortium* (2018).

[21] 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.

[22] Sarah Bird, Miro Dudík, Richard Edgar, Brandon Horn, Roman Lutz, Vanessa Milan, Mehrnoosh Sameki, Hanna Wallach, and Kathleen Walker. 2020. Fairlearn: A toolkit for assessing and improving fairness in AI. *Microsoft, Tech. Rep. MSR-TR-2020-32* (2020).

[23] 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.

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

[25] Virginia Braun and Victoria Clarke. 2019. Reflecting on reflexive thematic analysis. *Qualitative research in sport, exercise and health* 11, 4 (2019), 589–597.

[26] Robin N Brewer and Vaishnav Kameswaran. 2019. Understanding trust, transportation, and accessibility through ridesharing. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–11.

[27] Robin N Brewer and Anne Marie Piper. 2017. xPress: Rethinking design for aging and accessibility through an IVR blogging system. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 1–17.

[28] 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.

[29] Carrie J Cai, Emily Reif, Narayan Hegde, Jason Hipp, Been Kim, Daniel Smilkov, Martin Wattenberg, Fernanda Viegas, Greg S Corrado, Martin C Stumpe, et al. 2019. Human-centered tools for coping with imperfect algorithms during medical decision-making. In *Proceedings of the 2019 chi conference on human factors in computing systems*. 1–14.

[30] Tiffany Chan, Mara Mills, and Jentery Sayers. 2018. Optophonic reading, prototyping optophones. *Amodern* 8 (2018).

[31] Minsuk Chang, Stefania Druga, Alexander J Fiannaca, Pedro Vergani, Chinmay Kulkarni, Carrie J Cai, and Michael Terry. 2023. The prompt artists. In *Proceedings of the 15th Conference on Creativity and Cognition*. 75–87.

[32] Arindam Chaudhuri, Krupa Mandaviya, Pratixa Badelia, Soumya K Ghosh, Arindam Chaudhuri, Krupa Mandaviya, Pratixa Badelia, and Soumya K Ghosh. 2017. *Optical character recognition systems*. Springer.

[33] Tai-Yin Chiu, Yinan Zhao, and Danna Gurari. 2020. Assessing image quality issues for real-world problems. In *proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 3646–3656.

[34] Kimberle Crenshaw. 1989. Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics. In *University of Chicago Legal Forum*, Vol. 1989. 8.

[35] 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.

[36] Upol Ehsan, Samir Passi, Q Vera Liao, Larry Chan, I Lee, Michael Muller, Mark O Riedl, et al. 2021. The who in explainable ai: How ai background shapes perceptions of ai explanations. *arXiv preprint arXiv:2107.13509* (2021).

[37] Motahhare Eslami, Karrie Karahalios, Christian Sandvig, Kristen Vaccaro, Aimee Rickman, Kevin Hamilton, and Alex Kirlik. 2016. First I" like" it, then I hide it: Folk Theories of Social Feeds. In *Proceedings of the 2016 cHI conference on human factors in computing systems*. 2371–2382.

[38] Motahhare Eslami, Kristen Vaccaro, Min Kyung Lee, Amit Elazari Bar On, Eric Gilbert, and Karrie Karahalios. 2019. User attitudes towards algorithmic opacity and transparency in online reviewing platforms. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–14.

[39] Be My Eyes. Year. Announcing Be My AI. https://www.bemyeyes.com/blog/announcing-be-my-ai. Accessed: April 18, 2024.

[40] Philip M Ferguson and Emily Nusbaum. 2012. Disability studies: What is it and what difference does it make? *Research and Practice for Persons with Severe Disabilities* 37, 2 (2012), 70–80.

[41] Lisa Ferris. 2022. *Beyond the Hype: Getting Down to Business with Aira Visual Interpreter for the Blind.* https://lisaferris.medium.com/beyond-the-hype-getting-down-to-business-with-aira-visual-interpreter-for-the-blind-1c570dd318c3

[42] Martha Albertson Fineman. 2010. The vulnerable subject: Anchoring equality in the human condition. In *Transcending the boundaries of law*. Routledge-Cavendish, 177–191.

[43] Vinitha Gadiraju, Shaun Kane, Sunipa Dev, Alex Taylor, Ding Wang, Emily Denton, and Robin Brewer. 2023. " I wouldn't say offensive but...": Disability-Centered Perspectives on Large Language Models. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. 205–216.

[44] 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.

[45] Rosemarie Garland-Thomson. 1994. Redrawing the Boundaries of Feminist Disability Studies. *Feminist Studies* 20, 3 (1994), 583.

[46] Rosemarie Garland-Thomson. 2002. Integrating disability, transforming feminist theory. In *Feminist Theory Reader*. Routledge, 181–191.

[47] Rosemarie Garland-Thomson. 2011. Misfits: A feminist materialist disability concept. *Hypatia* 26, 3 (2011), 591–609.

[48] Leilani H Gilpin, David Bau, Ben Z Yuan, Ayesha Bajwa, Michael Specter, and Lalana Kagal. 2018. Explaining explanations: An overview of interpretability of machine learning. In *2018 IEEE 5th International Conference on data science and advanced analytics (DSAA)*. IEEE, 80–89.

[49] 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.

[50] Ricardo Gonzalez, Jazmin Collins, Shiri Azenkot, and Cynthia Bennett. 2024. Investigating Use Cases of AI-Powered Scene Description Applications for Blind and Low Vision People. *arXiv preprint arXiv:2403.15604* (2024).

[51] Naman Goyal, Cynthia Gao, Vishrav Chaudhary, Peng-Jen Chen, Guillaume Wenzek, Da Ju, Sanjana Krishnan, Marc'Aurelio Ranzato, Francisco Guzmán, and Angela Fan. 2022. The flores-101 evaluation benchmark for low-resource and multilingual machine translation. *Transactions of the Association for Computational Linguistics* 10 (2022), 522–538.

[52] Anhong Guo, Ece Kamar, Jennifer Wortman Vaughan, Hanna Wallach, and Meredith Ringel Morris. 2020. Toward fairness in AI for people with disabilities SBG@ a research roadmap. *ACM SIGACCESS accessibility and computing* 125 (2020), 1–1.

[53] Danna Gurari and Kristen Grauman. 2017. Crowdverge: Predicting if people will agree on the answer to a visual question. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 3511–3522.

[54] 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.

[55] Aimi Hamraie and Kelly Fritsch. 2019. Crip technoscience manifesto. *Catalyst: Feminism, theory, technoscience* 5, 1 (2019), 1–33.

[56] Jim Handy. 2012. Think Yelp Is Unbiased? Think Again. *Forbes* (16 August 2012). https://www.forbes.com/sites/jimhandy/2012/08/16/think-yelp-is-unbiased-think-again/?sh=24033ac711d1 Accessed: April 18, 2024.

[57] Donna Haraway. 2016. 'Situated Knowledges: the Science Question in Feminism and the Privilege of Partial Perspective'. In *Space, gender, knowledge: Feminist readings*. Routledge, 53–72.

[58] Christina N Harrington, Aashaka Desai, Aaleyah Lewis, Sanika Moharana, Anne Spencer Ross, and Jennifer Mankoff. 2023. Working at the intersection of race, disability and accessibility. In *proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–18.

[59] Alex Hern. 2020. Twitter apologises for 'racist' image cropping algorithm. *The Guardian* (21 September 2020). https://www.theguardian.com/technology/2020/sep/21/twitter-apologises-for-racist-image-cropping-algorithm Accessed: April 18, 2024.

[60] 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.

[61] Anna Lauren Hoffmann. 2021. Even when you are a solution you are a problem: An uncomfortable reflection on feminist data ethics. *Global Perspectives* 2, 1 (2021), 21335.

[62] Megan Hofmann, Devva Kasnitz, Jennifer Mankoff, and Cynthia L Bennett. 2020. Living disability theory: Reflections on access, research, and design. In *Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility*. 1–13.

[63] 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.

[64] Stacy Hsueh, Beatrice Vincenzi, Akshata Murdeshwar, and Marianela Ciolfi Felice. 2023. Cripping Data Visualizations: Crip Technoscience as a Critical Lens for Designing Digital Access. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–16.

[65] Jeremy Zhengqi Huang, Hriday Chhabria, and Dhruv Jain. 2023. "Not There Yet": Feasibility and Challenges of Mobile Sound Recognition to Support Deaf and Hard-of-Hearing People. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–14.

[66] 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.

[67] Caroline A Jones. 2017. *The global work of art: World's fairs, biennials, and the aesthetics of experience*. University of Chicago Press.

[68] Hernisa Kacorri, Kris M Kitani, Jeffrey P Bigham, and Chieko Asakawa. 2017. People with visual impairment training personal object recognizers: Feasibility and challenges. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 5839–5849.

[69] Rie Kamikubo, Kyungjun Lee, and Hernisa Kacorri. 2023. Contributing to Accessibility Datasets: Reflections on Sharing Study Data by Blind People. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–18.

[70] Mark T Keane, Eoin M Kenny, Eoin Delaney, and Barry Smyth. 2021. If only we had better counterfactual explanations: Five key deficits to rectify in the evaluation of counterfactual xai techniques. *arXiv preprint arXiv:2103.01035* (2021).

[71] Hyung Nam Kim. 2022. User experience of assistive apps among people with visual impairment. *Technology and Disability* 34, 3 (2022), 165–174.

[72] 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.

[73] Anja Kintsch and Rogerio DePaula. 2002. A framework for the adoption of assistive technology. *SWAAAC 2002: Supporting learning through assistive technology* 3 (2002), 1–10.

[74] Eva Feder Kittay. 1999. *Love's Labor: Essays on Women, Equality and Dependency* (1st ed.). Routledge.

[75] Vivian Lai and Chenhao Tan. 2019. On human predictions with explanations and predictions of machine learning models: A case study on deception detection. In *Proceedings of the conference on fairness, accountability, and transparency*. 29–38.

[76] Min Kyung Lee, Nina Grgić-Hlača, Michael Carl Tschantz, Reuben Binns, Adrian Weller, Michelle Carney, and Kori Inkpen. 2020. Human-centered approaches to fair and responsible AI. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–8.

[77] Talila A Lewis. 2022. Working definition of ableism–January 2022 update. *Talila A. Lewis* 1 (2022).

[78] Zachary C Lipton. 2018. The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue* 16, 3 (2018), 31–57.

[79] Henrietta Lyons, Eduardo Velloso, and Tim Miller. 2021. Conceptualising contestability: Perspectives on contesting algorithmic decisions. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–25.

[80] Kelly Mack, Emma McDonnell, Dhruv Jain, Lucy Lu Wang, Jon E. Froehlich, and Leah Findlater. 2021. What do we mean by "accessibility research"? A literature survey of accessibility papers in CHI and ASSETS from 1994 to 2019. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–18.

[81] 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.

[82] Tambiama Madiega. 2021. Artificial intelligence act. *European Parliament: European Parliamentary Research Service* (2021).

[83] Akihiro Maehigashi, Yosuke Fukuchi, and Seiji Yamada. 2023. Experimental Investigation of Human Acceptance of AI Suggestions with Heatmap and Pointing-based XAI. In *Proceedings of the 11th International Conference on Human-Agent Interaction*. 291–298.

[84] 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.

[85] 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).

[86] Emma J McDonnell, Soo Hyun Moon, Lucy Jiang, Steven M Goodman, Raja Kushalnagar, Jon E Froehlich, and Leah Findlater. 2023. "Easier or Harder, Depending on Who the Hearing Person Is": Codesigning Videoconferencing Tools for Small Groups with Mixed Hearing Status. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–15.

[87] MELT (Ren Loren Britton and Iz Paehr). 2023. ACCESS SERVER: Dreaming, practicing and making access. *First Monday* 28, 1 (2023). https://doi.org/10.5210/fm.v28i1.12908 Original work published January 16, 2023.

[88] Tim Miller. 2019. Explanation in artificial intelligence: Insights from the social sciences. *Artificial intelligence* 267 (2019), 1–38.

[89] Mara Mills. 2015. Optophones and musical print. (2015).

[90] Mia Mingus. 2017. Access intimacy, interdependence and disability justice. *Leaving evidence* 12 (2017).

[91] Julie Avril Minich. 2016. Enabling whom? Critical disability studies now. *Lateral* 5, 1 (2016).

[92] 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.

[93] Denis Newman-Griffis, Jessica Sage Rauchberg, Rahaf Alharbi, Louise Hickman, and Harry Hochheiser. 2023. Definition drives design: Disability models and mechanisms of bias in AI technologies. *First Monday* (2023).

[94] Joan Nwatu, Oana Ignat, and Rada Mihalcea. 2023. Bridging the digital divide: Performance variation across socio-economic factors in vision-language models. *arXiv preprint arXiv:2311.05746* (2023).

[95] Cecilia Panigutti, Ronan Hamon, Isabelle Hupont, David Fernandez Llorca, Delia Fano Yela, Henrik Junklewitz, Salvatore Scalzo, Gabriele Mazzini, Ignacio Sanchez, Josep Soler Garrido, et al. 2023. The role of explainable AI in the context of the AI Act. In *Proceedings of the 2023 ACM conference on fairness, accountability, and transparency*. 1139–1150.

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

[97] Betsy Phillips and Hongxin Zhao. 1993. Predictors of assistive technology abandonment. *Assistive technology* 5, 1 (1993), 36–45.

[98] Elisa Ramil Brick, Vanesa Caballero Alonso, Conor O'Brien, Sheron Tong, Emilie Tavernier, Amit Parekh, Angus Addlesee, and Oliver Lemon. 2021. Am i allergic to this? assisting sight impaired people in the kitchen. In *Proceedings of the 2021 International Conference on Multimodal Interaction*. 92–102.

[99] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. " Why should i trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 1135–1144.

[100] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2018. Anchors: High-precision model-agnostic explanations. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 32.

[101] Maria Riveiro and Serge Thill. 2021. "That's (not) the output I expected!" On the role of end user expectations in creating explanations of AI systems. *Artificial Intelligence* 298 (2021), 103507.

[102] Aja Romano. 2019. A group of YouTubers is trying to prove the site systematically demonetizes queer content. *Vox* (10 October 2019). https://www.vox.com/culture/2019/10/10/20893258/youtube-lgbtq-censorship-demonetization-nerd-city-algorithm-report Accessed: April 18, 2024.

[103] Francesca Rossi. 2018. Building trust in artificial intelligence. *Journal of international affairs* 72, 1 (2018), 127–134.

[104] Emma Sadjo, Leah Findlater, and Abigale Stangl. 2021. Landscape analysis of commercial visual assistance technologies. In *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility*. 1–4.

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

[106] Sandy's View. 2016. *How Can People Who Are Blind or Visually Impaired Tell Different Things Apart?* https://sandysview1.wordpress.com/2016/08/25/how-can-people-who-are-blind-or-visually-impaired-tell-different-things-apart/

[107] Jackie Leach Scully. 2008. *Disability bioethics: Moral bodies, moral difference*. Rowman & Littlefield.

[108] Shreya Shankar, Yoni Halpern, Eric Breck, James Atwood, Jimbo Wilson, and D Sculley. 2017. No classification without representation: Assessing geodiversity issues in open data sets for the developing world. *arXiv preprint arXiv:1711.08536* (2017).

[109] Sam Shead. 2020. Twitter to investigate racial bias in picture-cropping algorithm. *NBC News* (21 September 2020). https://www.nbcnews.com/tech/tech-news/twitter-investigate-racial-bias-picture-cropping-algorithm-rcna130 Accessed: April 18, 2024.

[110] Renee Shelby, Shalaleh Rismani, Kathryn Henne, Ajung Moon, Negar Rostamzadeh, Paul Nicholas, YILLA-AKBARI N'MAH, Jess Gallegos, Andrew Smart, and Gurleen Virk. 2022. Identifying sociotechnical harms of algorithmic systems: Scoping a taxonomy for harm reduction. *arXiv preprint arXiv:2210.05791* (2022).

[111] Hong Shen, Alicia DeVos, Motahhare Eslami, and Kenneth Holstein. 2021. Everyday algorithm auditing: Understanding the power of everyday users in surfacing harmful algorithmic behaviors. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–29.

[112] Margrit Shildrick. 1998. The Rejected Body: Feminist Philosophical Reflections on Disability, by Susan Wendell; Feminist Approaches to Bioethics: Theoretical Refections and Practical Applications, by Rosemarie Tong. *Women's Philosophy Review* 18 (1998), 68–72.

[113] Kristen Shinohara and Josh Tenenberg. 2007. Observing Sara: a case study of a blind person's interactions with technology. In *Proceedings of the 9th international ACM SIGACCESS conference on Computers and accessibility*. 171–178.

[114] Vivswan Shitole, Fuxin Li, Minsuk Kahng, Prasad Tadepalli, and Alan Fern. 2021. One explanation is not enough: structured attention graphs for image classification. *Advances in Neural Information Processing Systems* 34 (2021), 11352–11363.

[115] Tobin Siebers. 2008. *Disability theory*. University of Michigan Press.

[116] Ellen Simpson and Bryan Semaan. 2021. For you, or for" you"? Everyday LGBTQ+ encounters with TikTok. *Proceedings of the ACM on human-computer interaction* 4, CSCW3 (2021), 1–34.

[117] 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.

[118] 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.

[119] Pierre Stock and Moustapha Cisse. 2018. Convnets and imagenet beyond accuracy: Understanding mistakes and uncovering biases. In *Proceedings of the European conference on computer vision (ECCV)*. 498–512.

[120] Erik Štrumbelj and Igor Kononenko. 2014. Explaining prediction models and individual predictions with feature contributions. *Knowledge and information systems* 41 (2014), 647–665.

[121] Lucy Suchman. 2020. Agencies in technology design: Feminist reconfigurations. In *Machine ethics and robot ethics*. Routledge, 361–375.

[122] 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.

[123] Keeanga-Yamahtta Taylor. 2019. *Race for profit: How banks and the real estate industry undermined black homeownership*. UNC Press Books.

[124] Lida Theodorou, Daniela Massiceti, Luisa Zintgraf, Simone Stumpf, Cecily Morrison, Edward Cutrell, Matthew Tobias Harris, and Katja Hofmann. 2021. Disability-first dataset creation: Lessons from constructing a dataset for teachable object recognition with blind and low vision data collectors. In *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility*. 1–12.

[125] Anja Thieme, Ed Cutrell, Cecily Morrison, Alex Taylor, and Abigail Sellen. 2020. Interpretability as a dynamic of human-AI interaction. *Interactions* 27, 5 (2020), 40–45.

[126] Tovah. 2016. Sins Invalid's "Birthing, Dying, Becoming Crip Wisdom" Features Crip Art, Activism, Love & Liberation. *Autostraddle* (2016). https://www.autostraddle.com/sins-invalids-birthing-dying-becoming-crip-wisdom-features-crip-art-activism-love-and-liberation-354749/

[127] 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.

[128] Meredith Whittaker, Meryl Alper, Cynthia L Bennett, Sara Hendren, Liz Kaziunas, Mara Mills, Meredith Ringel Morris, Joy Rankin, Emily Rogers, Marcel Salas, et al. 2019. Disability, bias, and AI. *AI Now Institute* 8 (2019).

[129] Rua M Williams. 2021. I, Misfit: Empty Fortresses, Social Robots, and Peculiar Relations in Autism Research. *Techne: Research in Philosophy & Technology* 25, 3 (2021).

[130] Rua M Williams, Kathryn Ringland, Amelia Gibson, Mahender Mandala, Arne Maibaum, and Tiago Guerreiro. 2021. Articulations toward a crip HCI. *Interactions* 28, 3 (2021), 28–37.

[131] Bess Williamson. 2019. Accessible America: A history of disability and design. In *Accessible America*. New York University Press.

[132] Jacob O Wobbrock, Shaun K Kane, Krzysztof Z Gajos, Susumu Harada, and Jon Froehlich. 2011. Ability-based design: Concept, principles and examples. *ACM*

Transactions on Accessible Computing (TACCESS) 3, 3 (2011), 1–27.
[133] Austin P Wright, Zijie J Wang, Haekyu Park, Grace Guo, Fabian Sperrle, Mennatallah El-Assady, Alex Endert, Daniel Keim, and Duen Horng Chau. 2020. A comparative analysis of industry human-AI interaction guidelines. *arXiv preprint arXiv:2010.11761* (2020).
[134] Shaomei Wu, Jeffrey Wieland, Omid Farivar, and Julie Schiller. 2017. Automatic alt-text: Computer-generated image descriptions for blind users on a social network service. In *proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*. 1180–1192.
[135] Rui Zhang, Nathan J McNeese, Guo Freeman, and Geoff Musick. 2021. " An ideal human" expectations of AI teammates in human-AI teaming. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3 (2021), 1–25.
[136] Yunfeng Zhang, Q Vera Liao, and Rachel KE Bellamy. 2020. Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*. 295–305.
[137] 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.
[138] Yuhang Zhao, Shaomei Wu, Lindsay Reynolds, and Shiri Azenkot. 2018. A face recognition application for people with visual impairments: Understanding use beyond the lab. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–14.

A Interview Protocol

Note to readers: the following questions are merely an outline of the major topics we hoped to discuss with participants. Given the flexibility granted by semi-structured interviewing, we often deviated from this protocol, asked follow-up questions based on the specific stories that our participants shared, and tried to mimic participants' language as much as possible.

A.1 Current VAT use:

(1) You mentioned in the recruitment survey that you use [insert types of VAT]. Are there any other applications you would like to add or remove to this list?
(2) In general, how do you use [AI VAT]?
(3) If you could give advice to a blind person who is just starting to use these computer/AI-enabled applications what would you say?

A.2 Pre-interview scenarios

In our email exchange, you shared with me three examples of using [AI VAT]. I am going to go over these examples and ask a couple of questions.

A.2.1 Example 1: Low confidence of AI. For the first example, we asked you to share a scenario where you were confident that the visual assistance technology provided the wrong responses. You sent us [briefly describe this photo or read text] and [AI VAT] provided a response of [read].

(1) Can you tell me more about this example? Would you typically use [insert AI VAT name] for this task?
(2) What makes you confident that [insert AI VAT name] had the wrong output?
(3) Can you tell me your best guesses for why [insert AI VAT name] produced this particular result?
(4) Would you say other applications like [mention other AI VAT participant uses] could provide more accurate results?
(5) How do incidents like this, where VAT produced wrong responses, shape your experience of using this VAT in the future?
(6) In general, are there particular types of visual information or scenarios where you feel that VAT would fail to provide accurate responses? Can you share an example?
(7) If any, how do you address cases when AI VAT produced an error?

A.2.2 Example 2: Medium confidence/unsure of AI. For the second example, you shared a scenario where you were unsure if the visual assistance technology provided an accurate response. As a reminder, [briefly describe this photo or recap text] and [AI VAT] provided responses of [read].

(1) Can you tell me more about this example? What makes you unsure of [insert AI VAT name] output?
(2) How did you resolve this uncertainty?
(3) If applicable: [read the response of AI VAT]. What do you think of this response? How does the phrasing of this part shape your confidence about its quality?
(4) Beyond [discussed example], are there particular types of visual information or objects, that when using VAT, you often find yourself unsure of the credibility of its response? If so, tell me a recent example.
(5) Can you tell me about a time when you double-checked information from [insert name of AI VAT] using alternative sources maybe like another application, friend, or family member? What was the outcome?

A.2.3 Example 3: High confidence of AI. So for the first example, you shared a picture where you were confident that the visual assistance technology provided correct responses. As a reminder, [briefly describe photo or read text] and [VAT] provided the response [read].

(1) Can you tell me more about this example?
(2) What makes you confident or sure of [insert name of VAT] output in this example?
(3) Can you give me other recent examples where you were certain that [VAT] produced accurate information?
(4) Can you tell me about a time when you were confident about the response you received from VAT, but you later learned that it might have been inaccurate or wrong?

A.3 If the participant uses human-VAT: error in human VAT

So far we've discussed your confidence in AI-generated responses. I'm going to ask you a couple of questions about your experience in human VAT.

(1) Can you tell me about a time when you requested visual assistance from a volunteer/agent and you were uncertain about their response?
(2) How often does this happen?
(3) How is your process for accessing response quality different from when using humans vs. AI?

A.4 Future technologies

(1) In your opinion, what types of technology features AI VAT could introduce to help you better assess the quality or credibility of AI responses?
(2) If the creators of AI VAT could explain its general process for how it produces responses, so for example, VAT would

accessibly explain how their system works at a high level, do you think this may help you in understanding the quality of its particular response? Why or why not?
(3) If any, can you share an example of a situation where having an explanation for how a VAT response is generated would have been particularly useful to you in understanding its accuracy or credibility of AI VAT?
(4) If any, what potential risks or harms could arise from these explanations?

B Participants Overview

P#	Self-identified Technical Background
P1	Can solve most issues with some help from friends or family members or online
P2	Can solve most issues but then ask help from friends or family members or online
P3	Could easily solve most or all issues you encounter by yourself
P4	Can solve most issues with some help from friends or family members or online
P5	Can solve most issues with some help from friends or family members or online
P6	Seek professional help to fix technical issues
P7	Could easily solve most or all issues you encounter by yourself
P8	Can solve most issues with some help from friends or family members or online
P9	Could easily solve most or all issues you encounter by yourself
P10	Can solve most issues with some help from friends or family members or online
P11	Could easily solve most or all issues you encounter by yourself
P12	Could easily solve most or all issues you encounter by yourself
P13	Could easily solve most or all issues you encounter by yourself
P14	Can solve most issues with some help from friends or family members or online
P15	Could easily solve most or all issues you encounter by yourself
P16	Can solve most issues with some help from friends or family members or online
P17	Can solve most issues with some help from friends or family members or online
P18	Could easily solve most or all issues you encounter by yourself
P19	Can solve most issues with some help from friends or family members or online
P20	Ask friends and family members to fix issues
P21	Can solve most issues with some help from friends or family members or online
P22	Can solve most issues with some help from friends or family members or online
P23	Could easily solve most or all issues you encounter by yourself
P24	Could easily solve most or all issues you encounter by yourself
P25	Can solve most issues with some help from friends or family members or online
P26	Could easily solve most or all issues you encounter by yourself

Table 2: Participants' response to the recruitment survey question of "When it comes to solving technical issues, you often:" with options: 1) "Could easily solve most or all issues you encounter by yourself," 2) "Can solve most issues with some help from friends or family members or online," 3) "Ask friends and family members to fix issues,", and 4) "Seek professional help to fix technical issues." P2 clarified their response during our interview.