"I look at it as the king of knowledge": How Blind People Use and Understand Generative AI Tools

Rudaiba Adnin
Northeastern University
Boston, Massachusetts, USA
adnin.r@northeastern.edu

Maitraye Das
Northeastern University
Boston, Massachusetts, USA
ma.das@northeastern.edu

ABSTRACT

The proliferation of Generative Artificial Intelligence (GenAI) tools has brought a critical shift in how people approach information retrieval and content creation in diverse contexts. Yet, we have limited understanding of how blind people use and make sense of GenAI systems. To bridge this gap, we report findings from interviews with 19 blind individuals who incorporate mainstream GenAI tools like ChatGPT and Be My AI in their everyday practices. Our findings reveal how blind users navigate accessibility issues, inaccuracies, hallucinations, and idiosyncracies associated with GenAI and develop interesting (but often flawed) mental models of how these tools work. We discuss key considerations for rethinking access and information verification in GenAI tools, unpacking erroneous mental models among blind users, and reconciling harms and benefits of GenAI from an accessibility perspective.

CCS CONCEPTS

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

KEYWORDS

Accessibility, blind, visual impairment, Generative AI, ChatGPT

ACM Reference Format:

Rudaiba Adnin and Maitraye Das. 2024. "I look at it as the king of knowledge": How Blind People Use and Understand Generative AI Tools. 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, 14 pages. https://doi.org/10.1145/3663548.3675631

1 INTRODUCTION

The surge in Generative Artificial Intelligence (GenAI) tools reflects a broader paradigm shift in workflow and productivity. Nowadays, people are incorporating GenAI tools (e.g., ChatGPT [73], Google Gemini [33], Microsoft Copilot [65], and Claude [5]) into a wide variety of domains, including education [36, 62, 83], programming [16, 54, 92], communication [14], and content creation [45, 56]. Blind people are no exception to this. Accessibility technologies such as Be My Eyes and Envision have incorporated GenAI capabilities to assist blind users by answering visual questions [2, 87]. However, despite significant commercial and public attention towards the promises of GenAI for people with disabilities [37, 38], we know considerably less about how blind people use GenAI tools in their regular work and how they make sense of these tools for their needs. At this critical juncture of AI-driven world, understanding the use of GenAI among blind individuals—a large group that still remains underrepresented in many professions and higher education [7, 72]—is imperative to ensure they receive equitable opportunities to leverage the benefits of these emerging technologies.

While the prospects of GenAI are immense, so are the potential harms it can perpetuate, especially for people who are unaware of the risks associated with these technologies [13, 95]. An important way to mitigate these harms is to improve people's mental models of GenAI [98] so that they can question its capabilities and limitations and accordingly decide when and how to use these tools [55]. Given the complex and opaque nature of GenAI [24, 55] and in the absence of technical know-how, non-expert users run the risk of developing erroneous mental models and unrealistic expectations that are not consistent with the actual functionalities of these tools [98]. Although these risks apply to all users, the effect could be magnified for blind people, since an incorrect understanding of GenAI may reinforce and amplify the accessibility challenges [19, 52, 89], misinformation propagation [81], undue trust in unverified information [61], ableist biases [25, 60] and other technology-related harms [18, 96] blind people already encounter. Thus, to better understand GenAI accessibility for blind people, we investigate: *How do blind people use GenAI tools and for what purposes? How do they navigate challenges and biases, if at all, while using GenAI? What mental models do they develop to make sense of GenAI tools?*

To this end, we present findings from interviews with 19 blind individuals who have experience with GenAI chatbots such as ChatGPT, Copilot, Gemini, and Claude and GenAI-powered image description tool Be My AI (a feature of the Be My Eyes app). Our analysis shows that blind individuals use GenAI tools for various content creation and information retrieval tasks, while navigating critical accessibility issues on GenAI interfaces and working through the inaccuracies, hallucinations, and idiosyncracies of GenAI responses. We also detail the ways in which blind individuals form—at times flawed and oversimplified—mental models of GenAI tools. Finally, we highlight how blind users grapple with concerns about ableist biases and other harms perpetuated by GenAI tools against the benefits they receive from using these tools.

Overall, our paper makes three key contributions. First, we present rich empirical understandings of how blind people integrate GenAI tools to enhance productivity and access to information in their regular work practices, extending prior research that investigated disability representation and biases in GenAI [25, 29, 60] and potentials of GenAI to support access needs of people with

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

ACM ISBN 979-8-4007-0677-6/24/10.
https://doi.org/10.1145/3663548.3675631

neurodivergence [32, 46, 86] and other disabilities like chronic illness and aphantasia [30]. Second, we unpack how blind people negotiate the consequences of inaccurate GenAI content and the effort needed to verify information by reasoning through competing factors, such as context, stakes, verifiabililty, and believability. Finally, we bring forth one of the first detailed accounts of blind individuals' mental models of GenAI as shaped by their assumptions of these tools' work processes, sources of information, and response generation approach. We revisit the similarities and differences of these mental models with that of sighted users [98] to surface how (in)accessibility of GenAI tools uniquely shape blind people's perceptions of GenAI capabilities and limitations, opening up future research directions in AI and accessibility.

2 RELATED WORK

We situate the present study within prior research on generative AI (GenAI) tools and practices, the intersection of AI and accessibility, and mental models of AI systems.

2.1 Generative AI Tools and Practices

With the proliferation of large language models (LLMs), GenAI tools like ChatGPT, Copilot, and Gemini have gained immense popularity across various domains, including education [36, 49, 83, 99], programming [16, 54, 92], communication [14], and creative work [45, 56, 84]. Scholars have started exploring how to better support users' interaction with GenAI to enhance their productivity [45, 92]. To this end, researchers have built new systems e.g., collaborative design application [88], conversational game [14], and tools for writing better prompts [9, 93]. Others have investigated how users incorporate GenAI into their workflows. For instance, prior work found that students are interested in using GenAI for brainstorming new ideas [62] and addressing coursework-related queries [11]. These tools can also help prepare solutions to programming problems [4, 16, 74, 99] and completing coding tasks quickly [54]. However, developers tend to avoid using AI assistants due to the challenges in controlling these tools to produce desired output [54]. In the creative domain, Inie et al. [45] found that creative professionals are concerned about intellectual property issues and GenAI weakening human creative sparks. Furthermore, GenAI often produce inaccurate and outdated information [11, 99] and fabricated but plausible-sounding content, commonly known as hallucinations [15], which limit the reliability of these tools in important use contexts. Due to these challenges, users often do not trust GenAI output and feel the need for human supervision of AI-generated answers [4].

2.2 Research on AI and Accessibility

Set against the large and growing literature on GenAI tools and practices, limited studies have explored the implications for GenAI among people with disabilities. Notably, Glazko et al. [30] conducted an authoethnographic study within a team of researchers with and without disabilities to demonstrate how they used GenAI to create access for themselves and others and how existing GenAI tools sometimes failed in this regard. Others examined GenAI use among AAC users [86] and autistic people [46], highlighting that GenAI can save time and reduce physical and cognitive effort during communication, but these tools need to reflect users' communication preferences [86]. Researchers also investigated disability representation in GenAI and found ableist biases and stereotypes in GenAI responses [25, 29, 60]. Focusing specifically on GenAI accessibility for blind people, Das et al. [18] identified image provenance (i.e., information about image source) and aberrations (i.e., unrealistic depictions in images) as important information desired by blind users in the description of AI-generated images. Relatedly, Huh et al. [43] built a system to make text-to-image generation more accessible for blind users by providing detailed descriptions of the AI-generated images and allowing options to verify if generated images follow their prompts.

While research on GenAI use among blind people remains nascent, a larger body of work examines blind people's experience with other AI technologies. Researchers incorporated teachable AI to assist blind people in finding their personal objects [39, 68] and detailed what factors blind people assess when sharing their information for AI datasets [47]. Others uncovered accessibility benefits and challenges blind users experience while using voice assistants [1, 76]. Collectively, prior work shows the many potential benefits, harms, and considerations of AI for accessibility. Situated in this literature, our study contributes to a detailed understanding of how blind people use and understand GenAI tools in their regular work and the challenges and opportunities these tools present for their practices.

2.3 Mental Models of AI Systems

Studying accessibility of complex and opaque systems like GenAI requires unpacking how users form mental models i.e., conceptual representations of the systems based on their experience interacting with those systems [50, 71]. Without a clear conceptual understanding, individuals often form their own simplified mental models of how a system works that do not always correspond to the system's actual functionalities [71]. Over the years, HCI scholars have investigated users' mental models of complex systems including AI technologies [8, 26, 42, 66]. Researchers found that users with oversimplified mental models of voice assistants have a limited understanding of the privacy risks associated with those tools [42]. When users encounter unexpected behaviors from voice assistants, they require explanations to refine their mental models for more effective interaction [40]. Others captured how individuals develop mental models of the error boundary of AI systems, highlighting that a good mental model can assist individuals in achieving better performance [6]. To empower individuals to actively engage with AI tools rather than being passive consumers and to facilitate informed decision-making, researchers call for increased effort to promote public AI literacy [58, 85].

Recently, researchers have started exploring how individuals conceptualize the responses from GenAI tools [3, 97]. Liao and Wortman Vaughan [55] caution that interacting with LLMs with flawed mental models can lead to unsafe use, over-reliance, and other interaction-based harms. Further, oversimplified and erroneous mental models of LLMs encourage disclosures of sensitive topics which leads to privacy risks [98]. Our study extends this scholarship that focused on non-disabled people by contributing to new insights about mental models of GenAI tools from the perspectives of blind individuals.

3 METHOD

3.1 Participants

We conducted interviews with 19 blind and visually-impaired individuals who had experience using GenAI chatbots such as ChatGPT 3.5 or 4, Google Gemini (formerly Bard), Microsoft Copilot (formerly Bing Chat),[1] and Claude and Gen-AI powered image describer, Be My AI (a feature of the Be My Eyes app). Participants were recruited using an online survey circulated through an organization that works with blind people, our research networks, and snowball sampling. Out of 27 respondents, we selected 19 participants, screening for their GenAI usage in the last four months. Most participants were intermediate users of GenAI except three frequent users and one beginner. Four participants reported using text-to-image tools like DALL-E and Midjourney a few times; however, we centered our focus on participants' experience with text-based GenAI chatbots like ChatGPT and image description apps like Be My AI. In addition to GenAI tools, all but one participant frequently used voice assistants (e.g., Amazon Alexa and Siri) and all but one used image description apps (e.g., Seeing AI, Google Lookout, and Envision AI). Participants primarily used one or more screen readers (e.g., JAWS, NVDA, and VoiceOver) for information access, although eight also used braille displays. All participants except two lived in the US. Table 1 shows details of participants' self-reported visual disabilities, occupation, GenAI tools used, and frequency of using GenAI. Table 2 shows participants' demographic information on an aggregate level to maintain anonymity.

3.2 Procedure

We conducted semi-structured interviews remotely over Zoom between January–March 2024, with approval from our university's Institutional Review Board. Interviews began with obtaining participants' verbal consent. We first asked participants to share what GenAI tools they used and for what purposes. We requested them to walk us through their process of interacting with their preferred text-based GenAI chatbot (e.g., ChatGPT, Copilot, Gemini) via screen sharing, along with sharing the screen reader speech. Participants showed examples from their previous chat histories and performed live demonstration of how they formulated prompts, read answers, and wrote follow-up prompts and any accessibility issues they encountered on these tools. To demonstrate Be My AI, participants opened the Be My Eyes app on their phone, since it was not available on desktop. Their phone screen reader read out the image description generated by Be My AI. We requested participants to increase their phone volume and bring it closer to their computer (which they were using for our interview) so that we could listen to and record the description through Zoom. We probed participants for deeper reflections on the descriptions generated by Be My AI. For this demonstration, we sent participants two sample images before the interviews: one showed a person with a dog and the other showed two persons in a shopping outlet. The images can be found at these links: image 1 and image 2. Some participants also used their own images and shared those with us after the sessions. Below we provide the description generated by Be My AI for the sample image with a dog.

> "The picture shows a man walking in the park with a guide dog. The man is holding a white cane in his left hand and the dog leash in his right hand. He's wearing sunglasses, a light jacket, a scarf, and casual trousers with sneakers. The dog appears to be a black Rottweiler with a red collar. They are on a dirt path surrounded by lush green trees and some grassy areas in the background. There is a small bridge over a stream with a person crossing it. The setting is peaceful and suggests a quiet and natural environment."
> Note that this Be My AI-generated description alternated the items in the person's hands. In the image, the white cane is in their right hand and the dog leash is in their left hand. Also, there is no visual sign that identifies the dog as a guide dog. It is not wearing any harness that guide dogs commonly wear.

To understand participants' mental models of GenAI, we drew on the five big ideas of AI [85] and asked participants to share their thoughts on how these tools worked, whether and how these tools could understand their questions, how the responses were generated, how they perceived the quality of the responses, and the overall capabilities, limitations, and social impacts of these tools. We probed participants with a particular emphasis on instances of inaccurate or unexpected responses, since expectation violation has been found to be helpful in revealing user mental models [20]. To keep our study procedure accessible, we did not incorporate any mental model drawing activity [48, 98]. Instead, we developed our interview protocol following prior studies that used interviews to reveal users' mental models of technologies [20, 22, 78]. We intentionally avoided technical jargon like LLM, training data, or 'generative AI' unless participants mentioned these terms themselves. We did not answer any questions from participants about how AI or GenAI worked. Interviews lasted for about 60-90 minutes. Participants were compensated with US$30 per hour (prorated) via Amazon gift card or PayPal. All interviews were video-recorded and transcribed for analysis.

3.3 Data Analysis

We analyzed data following a reflexive thematic analysis method [10]. Taking an inductive approach, the first author open-coded the entire corpus while both coauthors closely read and reviewed all codes and the data. Our initial codes captured instances, such as lack of keyboard navigation support, hallucinations, techniques for verifying accuracy, and more. We met weekly to discuss the codes and excerpts and compare data to data and data to codes to develop initial themes. Through this iterative process, we constructed five overarching themes that capture the core aspects of the ways in which blind individuals use and make sense of GenAI tools.

4 FINDINGS

Our analysis reveals that for blind people, using mainstream GenAI tools in everyday practices involves leveraging its strengths for content creation and information retrieval while navigating various accessibility issues, inaccuracies, and idiosyncracies of these tools. In doing so, blind people develop interesting (and at times erroneous) mental models about how GenAI work and think through the harms and biases associated with these technologies.

[1]While describing individual participants' experience, we use the name or version of the tool they reported.

Table 1: Details of interview participants. All names are pseudonyms. GenAI Usage: Frequency or an approximate count of times GenAI tools were used in the last 4 months prior to the interviews. Gemini (formerly Bard) is a Google product. Copilot (formerly Bing Chat) is a Microsoft product. We report the tool name/version participants mentioned. All participants regularly used Be My AI. * denotes that the participant has programming experience, although not everyone had expertise in AI.

Name	Self-reported Visual Disability	Occupation	GenAI Tools Used	GenAI Usage
Adam*	Totally blind	Accessibility experience designer	ChatGPT 4.0	5-6 times a day
Bella*	Totally blind	Adaptive tech instructor	ChatGPT 4.0, Gemini, Bing Chat	2-3 times a week
Carla	Legally blind with limited light perception	Clinical psychologist	ChatGPT 3.5	Once a week
Daisy	Totally blind	Personal care provider	Gemini, ChatGPT	>15 times
Ethan	Blind	Assistive tech manager	ChatGPT 3.5, Copilot	>15 times
Frank*	Retinal degeneration	Retired	Bing Chat, ChatGPT, Bard	>15 times
Gina	Blind	Instructor for the blind	ChatGPT 4.0, Gemini	>15 times
Henry*	Totally blind. No light perception	Musician	ChatGPT 3.5, Bard	>15 times
Ivan*	Blind since birth. Leber's Congenital Amaurosis	Adaptive tech instructor	ChatGPT 4.0, Bard	>15 times
Julia*	Visual impairment. Glaucoma, Retinopathy of Prematurity	Community manager	ChatGPT 3.5, Bard, Bing Chat, Claude	>15 times
Kevin	Blind with limited residual vision	Small business owner	ChatGPT 3.5, Claude	>15 times
Lily*	Blind	Accessibility tester	ChatGPT 3.5, Bard, Copilot	>15 times
Mike*	Totally blind. Some light perception	Student	ChatGPT 3.5, Bard, Claude	>15 times
Nancy	Totally blind	Works in a committee	ChatGPT 3.5, Copilot, Gemini, Perplexity	11-15 times
Noah*	Legally blind all life, totally blind 4+ years. Glaucoma, cataracts, and Corneal Edema	ADA compliance testing	ChatGPT 3.5	6-10 times
Portia*	Totally blind	Advocate	Claude, ChatGPT, Bard	6-10 times
Ruby	Totally blind 4+ years	Accessibility trainer intern	ChatGPT 3.5, Copilot	6-10 times
Sara*	No vision	Worked for tech support	ChatGPT 3.5, Bing Chat	6-10 times
Theo	Totally blind. Retina damage and cataract in right eye; prosthetic left eye	Unemployed	Bard	1-5 times

Table 2: Participants' (n=19) demographic information on an aggregate level

Gender	Count
Male	8
Female	10
Not disclosed	1

Age (years)	Count
18–24	1
25–34	8
35–44	3
45–54	2
55–75	3
Not disclosed	2

Race	Count
White	10
Black	1
Hispanic	2
British	1
Asian	1
Not disclosed	4

4.1 Adapting to Accessibility Issues in Generative AI Tools

Many widely adopted technologies are rife with accessibility issues that blind people must navigate by devising various workarounds and coping mechanisms [19, 59, 89]. GenAI tools are no exception to this; while on the surface text-based GenAI tools like ChatGPT may appear to be *"technically"* accessible, our participants reported encountering a number of challenges due to these tools' disregard for established Web Content Accessibility Guidelines (WCAG) [90] coupled with their *"terrible UI"* (Kevin).

During our sessions, all participants demonstrated that buttons for copying, regenerating, and downvoting ChatGPT responses were unlabeled (Figure 1). Hence, blind users must figure out the functionalities of these buttons through trial and error or ignore them altogether. Moreover, ChatGPT and Claude neither provided appropriate heading labels, regions, or landmarks for screen reader users to swiftly traverse around the interface nor enabled any shortcuts to jump between previous and next prompts or responses. Our

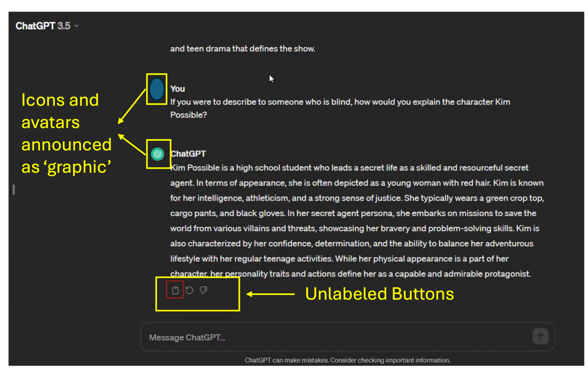

Figure 1: Screenshot of the ChatGPT 3.5 interface. Unlabeled buttons are marked, e.g., buttons for copying, regenerating, and downvoting responses. Screen readers announced the user's avatar and the ChatGPT icon near the prompt/response as 'graphic', and blind users repurposed the shortcut to move between graphics as a workaround for quick navigation.

participants had to spend considerable time in finding GenAI responses or navigating to the message box to type a new prompt, since it required them to *"brute force"* (Kevin) their way through the entire conversation repeatedly by scrolling with arrow keys. Participants who used Copilot and Bard also expressed frustration with circumventing extraneous sample prompts and ads, which created *"a lot of clutter to sort through"* (Ruby) before reaching the chat fields. Furthermore, blind users did not get any instantaneous notification when the system finished response generation, requiring them to *"dig for it"* (Portia) by moving screen reader focus.

To adapt to these challenges, our participants came up with various workarounds. Some repurposed the shortcut to move between graphics (G + Screen Reader Modifier) to quickly jump between prompts on ChatGPT, since the user's avatar and the ChatGPT icon adjacent to the prompt / response were announced as 'graphic' by the screen reader (Figure 1). Carla, Gina, and Frank used GenAI tools on phone because they found it relatively easier to locate information on a smaller screen using tap gestures. However, this approach *"still involved a fair amount of hide and seek… because sometimes the page scrolled"* (Frank) inconsistently. Kevin, in contrast, avoided reading AI responses on the native apps altogether and manually copied those to a notepad for reading and editing. Participants also noted several usability issues on Be My AI, for example, not being able to import images directly into the app from the gallery and losing conversation history once they exited the chat instance for a particular image. Although some of these difficulties (e.g., extraneous ads) may affect sighted users as well, our participants highlighted the compounding impact of navigating these challenges using screen readers, which *"slowed them down"* (Julia) and made their experience with GenAI tools *"annoying"* (Gina). Thus, unlike sighted users, blind individuals must work through additional accessibility issues to reap the benefits of GenAI tools.

4.2 Leveraging Generative AI for Content Creation and Information Retrieval

Our blind participants incorporated GenAI tools in a wide variety of content creation and information retrieval tasks, ranging from preparing copywriting materials, emails, course outlines, resumes, elevator pitch, cover letters, recommendation letters to programming and creative writing (e.g., stories, poems, songs). Although many of these use cases and the advantages of GenAI our participants described align with the experience of sighted users [45, 84], our analysis foregrounds certain contexts in which using GenAI tools carry important and unique implications for blind individuals, such as visual question answering [31, 35]. Below we broadly describe how our participants utilize GenAI tools to enhance their workflow, drawing out specific examples that relate to their experience with blindness.

Echoing findings from recent work involving sighted users [27, 28, 84], blind participants shared different strategies they adopted while using GenAI tools for creating content. In one strategy, participants started with one or more source material(s), such as an outdated resume, an article, or quick scribbles jotted during a meeting, and then *"ran it through the AI"* (Portia) to summarize, expand, combine, rephrase, or organize those materials into revised and improved content. In some cases, participants chained outputs from multiple GenAI tools to create the final product. For instance, Gina and Julia used Be My AI to produce descriptions of images, copied those into ChatGPT, and prompted ChatGPT to write social media posts or stories based on the image descriptions. In another strategy, participants used GenAI as a *"jumping off point"* (Ivan) for brainstorming different possibilities when they felt *"clueless about how to approach something"* (Mike). Julia and Bella described using GenAI to *"make [their] own content more accessible for others"*, for example, by creating accessible webpages or forms [30]. Many participants—especially those who were English language learners or were not proficient in writing—used GenAI to fix spelling, grammar, and formatting errors and translate text from one language to another. These participants felt that GenAI made writing tasks *"a lot less daunting"* (Julia) for them. Such proofreading support is crucial for blind people, given that blind screen reader users are more likely to make spelling [53, 75] and formatting errors due to accessibility issues in writing applications [19, 67].

In addition to content creation, most participants also used GenAI for information seeking [12, 41, 83], for instance, searching about TV shows, products for shopping, or accessibility guidelines. Some participants used GenAI for planning events or getting advice on handling everyday situations. As examples, Bella gathered ideas about tactile activities to throw a party for her blind daughter's birthday, Ethan queried suggestions about raising a child as a blind parent, Mike generated a roadmap on how to manage a PR (public relations) vertical for his college fest as a blind person, Ruby curated a weight loss program, and Daisy consolidated her health symptoms before talking to her doctor.

Related to information seeking, one unique use context of GenAI for blind individuals is visual question answering [31, 35], for which participants primarily used Be My AI but also sometimes newer GenAI models that can describe visual information e.g., ChatGPT 4. Almost every participant appreciated that compared to *"one or two sentence"* (Portia) descriptions given by sighted people, Be My AI provided richer descriptions in a systematic way, starting with the foreground followed by the background, including details of people's attire, surroundings, objects, colors, and the overall vibe. Gina shared, *"I've actually just grown used to the Be My AI descriptions because some people just don't know how to describe things to blind people. They have no idea what [blind people] can and cannot see and what they want to know and don't wanna know."*

Participants appreciated using GenAI for the above-mentioned tasks because it made their workflow *"efficient"* (Carla) such that they *"didn't have to spend hours"* (Gina) to search information online [12] and combine, revise, reformat, and proofread all that information to produce the end result. Additionally, participants felt that GenAI tools helped them develop new skills and enhance proficiency in areas they had tried to learn before but did not have much success. As self-learners, some participants found it helpful to receive feedback and explanations from GenAI when they needed pointers to get unstuck on problems. Julia shared, *"I don't feel silly asking stupid questions"* to GenAI. Among our participants, Nancy tried to learn songwriting with ChatGPT, Mike learned simplified explanations of academic jargon, and Bella explored mathematical concepts. Reflecting on how ChatGPT helped him practice programming, Henry said,

> *"I'm not very good at taking in heaps of information that come from all these [coding] documents. So, trying to teach myself program has been a nightmare. Then again, I've never been able to find people who have time or patience to teach me how to do it either... And I've learned more from ChatGPT in the past year than I've learned in the past 20 years... ChatGPT told me what functions I needed [to create an Windows app]. I was able to sort of look for those [functions]... in the relevant documentation... and look against the snippet of code it gave me and I was like, that should actually work."*

Despite the benefits elaborated above, participants acknowledged the limitations of AI-generated content in terms of linguistic quality. Most participants observed that ChatGPT presented content in a distinct *"bulleted"* structure, which made it appear *"too robotic"* (Ethan, Ivan, Daisy), *"very bookish"* (Mike), and *"formulaic"* (Daisy). They also noted that ChatGPT produced redundant phrases and overused certain *"huge words"* (Lily). Participants characterized this syntax as *"way over the top"* (Bella, Daisy), *"really flowery... almost too sweet"* (Ruby), and *"overly verbose in a way that doesn't quite feel human"* (Ivan). Most participants felt that they can easily recognize *"the ChatGPT style"* (Noah) when they encounter unknown text. Hence, while creating content, participants made sure to readjust GenAI responses to eliminate overused words and *"sprinkle in some of me"* (Daisy). Nancy reflected on this balancing act in human-AI creation: *"I want AI to help me out, but I also want to put in my own words… I don't want it to be 100% AI. So, I definitely modify it where it sounds good, but it's also coming from me."* Considering this multi-step process for reviewing and editing GenAI responses, Ethan refrained from using GenAI for tasks like writing emails *"because it's actually more work."* These examples highlight that incorporating GenAI in work practices often requires some extra effort from our participants. In some cases though, having to tune the language and structure of responses is just the start, as there is additional work that is necessary to navigate the inaccuracies and quirks of GenAI responses.

4.3 Working through Inaccuracies and Idiosyncrasies of Generative AI

A known limitation of mainstream GenAI tools is their tendency to generate information that is fabricated, inaccurate, or inconsistent with input data [15]. To tackle these inaccuracies and idiosyncracies, our participants need to navigate the challenges associated with verifying generated information and improving response quality.

4.3.1 Identifying Hallucinations and Verifying Accuracy. Our participants shared many examples of factually inaccurate or fabricated information (i.e., hallucination) provided by GenAI. For instance, Be My AI described Daisy's raincoat pattern as *"hearts and stars"* although it was *"clouds and raindrops"* and ChatGPT replaced Ethan's name with a fictitious name 'Chris' when he asked it to revise and update his old resume. Daisy emphasized that GenAI tools projecting *"confidence"* in their hallucinated responses, especially regarding visual information, *"can be misleading to someone who isn't able to visually verify for themselves what something looks like."* To assess the validity of GenAI responses, blind participants tried to sense whether those *"sounded really weird"* (Adam) or were *"very different"* (Lily) from what they had anticipated based on their prior knowledge. For instance, while exploring a sample image, Ruby, Sara, and Ethan doubted Be My AI's description of the dog's breed. Ruby said, *"Knowing what I know about guide dogs, Rottweilers aren't generally a breed that's used."* Refer to Section 3.2 for the full image description given by Be My AI.

When in doubt, participants tried to find alternate ways to validate GenAI responses. Most often, they *"turned to Google"* (Ivan) or visited websites that were likely to contain accurate information about their queries [84]. For example, Bella checked Freedom Scientific webpage for confirmation on a JAWS screen reader command. Besides checking external applications, participants sometimes used the GenAI tools to assess accuracy. They repeated the question to the same tool at different times (Carla) or by starting a new chat (Julia) to see if they received different responses. Alternatively, they ran it through multiple equivalent genAI tools (e.g., ChatGPT and Gemini) to confirm whether they got similar responses after repeated try. Additionally, participants engaged in a process of *"deductive reasoning"* (Ethan) with the GenAI tool by asking follow-up questions to judge the accuracy or completeness of its response. For instance, Ethan probed Be My AI about the breed of the guide dog in our sample image: *"You sure it's a Rottweiler?"* Upon receiving a confident response, he followed up, *"Prove to me that it's a guide dog."* In response, Be My AI acknowledged that it assumed the dog to be a guide dog due to the presence of a white cane in the image.

Given the back-and-forth process required for verification, our participants were judicious about when they must check accuracy of GenAI responses or when they could forego checking. They agreed it was not safe to *"100% rely on AI"* (Nancy) for information that would be used to make financial, medical, or health-related decisions (e.g., which products to buy, which medicines to take, or food expiry date), incorporated in a professional or academic context (e.g., writing a paper), or shared publicly (e.g., on someone's website). However, personal use cases had more *"tolerance for errors"* (Frank) where participants felt accuracy *"doesn't matter. It's not the end of the world"* (Ethan).

Nevertheless, we observed an overall trend among participants toward minimizing the gravity and likelihood of hallucinations. Several participants commented that they *"generally trust"* ChatGPT responses for *"high-level details"*, considering those to be right *"90%"* (Kevin, Ivan) or *"99.999% of the time"* (Adam) [61]. Their trust on GenAI responses bolstered when tools like Copilot or Perplexity cited links to source websites. Portia said, *"It'll already give me references to where that information came from, like according to the Journal of Psychiatry... So then I don't have to fact check."* Even those who were more skeptic about GenAI accuracy (Ivan, Lily, Ruby, Julia, Daisy, Mike) tended to trust image descriptions from Be My AI because those seemed to be *"detailed enough"* (Ruby) and also because participants appreciated the *"independence"* afforded by Be My AI descriptions over seeking sighted help for verification.

4.3.2 Enhancing Response Quality through Prompt Engineering. Blind participants assumed suboptimal prompts or other inputs to GenAI to be a likely reason for erroneous or low-quality responses. For instance, all participants' first reaction to inaccurate image descriptions from Be My AI was poor image quality or visually uncertain objects in input images. They commented that lighting, camera angle and distance from the target object, blurriness, reflection or glare, partially obscured objects—all these could *"confuse"* (Frank) GenAI tools and lead to inaccurate image descriptions. Some participants were more willing to critique the photos they took than attribute unsatisfactory image descriptions to GenAI limitations, and often tried to capture better photos before running it through GenAI. In doing so, they exhibited a tendency to downplay GenAI limitations while putting the onus of capturing good quality photos on themselves. For instance, while describing a sample image, Be My AI consistently alternated the items in the person's left and right hands (refer to Section 3.2 for the full description). To speculate why this might have happened, Sara said, *"People have told me when I've uploaded things...it'll stretch them out or do wacky stuff to my pictures. So, I wonder if it got turned around when it got uploaded,"* although this was not the case. Likewise, participants tended to blame the quality of their text prompts when interpreting possible reasons for *"off base"* GenAI responses. Ethan said, *"If I get a bad response, I gave it a bad prompt. It's my fault."*

Participants reflected on their effort to learn *"prompt crafting"* or *"prompt engineering"* to improve GenAI response quality [51]. For example, they often tried to create specific and detailed prompts *"to lead ChatGPT in the correct direction"* (Carla). Ethan, Julia, Henry, Adam, and Kevin considered follow-up questioning as a useful mechanism to remind GenAI tools their original asks if it started making assumptions about their queries. Julia explained: *"I just played whack-a-mole and whack-a-mole until it finally did what I wanted it to do. And you have to problem-solve these instances to realize … What did it assume? Let's figure that out and fix it."* Interestingly, other participants recalled that follow-up questioning degraded response quality, because the GenAI tools *"lost track of the conversation"* (Mike) as the chat got longer than *"5 or 10 messages"* (Henry). This divergence in opinions indicates how participants formed different ideas and expectations about GenAI based on their own experiences and perceptions about these tools.

4.4 Developing Mental Models of Generative AI

Our participants actively tried to hypothesize how GenAI tools worked and accordingly adapted their interactions with these tools. They described *"playing around"* with these tools by progressively asking simple to advanced questions to *"test its knowledge"* (Sara) and *"learn its limits"* (Adam). They also learned about GenAI tools from discussion with friends, family members, and other blind users, accessibility webinars and training programs, news articles, official documentations, and participating in beta testing of GenAI tools. Below we distill salient mental models of GenAI tools we observed among our participants. These mental models particularly relate to the text generation capabilities of GenAI and do not dive deeper into image recognition processes of tools like Be My AI. While some mental models align with that of sighted users [98], there are important differences that are shaped by our participants' experience with blindness and usage of assistive technologies, and we return to this point in the Discussion (Section 5.2).

4.4.1 "Google Search on Steroids". Most participants believed that GenAI tools *"pulled up information from the internet"* either by directly searching and/or using a search engine like Google or Bing under the hood. Ruby (ChatGPT 3.5 user) gave an example where she felt that ChatGPT response to her query about a weight loss program *"might have been pulled from maybe a nutritional website or medical website, the fact that it tells you to consult with your doctor."* Informed by this mental model, most participants used GenAI tools as a *"replacement of Googling"* (Ivan), given that it significantly streamlined their information foraging workflow [12]. Ethan called ChatGPT *"Google search on steroids"* that can *"deliver what you need to know right on the screen without having to sift through"* (Gina) links and articles returned by search engines. Certain GenAI limitations bolstered this perception. Participants speculated that GenAI occasionally provided inaccurate information because it gathered data from outdated and unverified internet content.

While some participants used GenAI tools that indeed had internet search capabilities (e.g., ChatGPT 4, Copilot), this perception was also evident among participants who used ChatGPT 3.5 that did not have such features. Some ChatGPT 3.5 users knew that it was unable to correctly answer questions about recent events. However, they were either unsure exactly how this older version collected information or assumed that it also gathered information from the internet. Nancy (ChatGPT 3.5 and Copilot user) said, *"If you are researching something on Copilot, I think it literally goes to Google and pulls up information. I think for things like ChatGPT, it probably does that too, but I don't know how it's different because it's not as current as Copilot."*

4.4.2 "King of Knowledge". Unlike the previous mental model that equated GenAI tools with a search engine, some participants speculated that GenAI tools conducted a keyword-based query directly into one (or more) of its *"massive database"* which contained *"all kinds of information"* gathered from textbooks, archives, coding manuals, and other web content. Portia elaborated, *"I feel like the information has to sit somewhere, even if it's in the cloud… It (ChatGPT 3.5) doesn't just spit out of nothing… It had to go have those answers from some database that has already researched it… I don't know how that database got created or who's adding or removing."* Due to this mental model, Sara believed in the superiority of ChatGPT for information seeking: *"Honestly, I look at it as the king of knowledge… So, if ChatGPT doesn't know what to say, I'm just not gonna find what I'm looking for because it ain't there."*

Interestingly, Ethan and Bella, who were familiar with the term LLM, thought of it as a huge database on which GenAI tools performed a keyword-based search. Ethan said,

> *"An LLM is basically an infinite amount of data essentially that's been fed into a computer. That computer can then parse for information that you're looking for. They can gather specific data from that model to give you an answer to something you wanted. So, if you ask a question, it goes back into its system and sees—Has that question been asked before? Can I string information together?—to give that person an answer."*

The above two mental models guided our participants' belief that GenAI tools were better at addressing *"straightforward"* (Noah) questions that had *"factual"* (Bella, Noah), *"concrete"* (Lily), and *"objective"* (Portia) answers, because the tools can *"come back with pretty much correct answers based on information it pulls"* (Bella).

4.4.3 *"Word Generating Machine".*

Unlike models described in Sections 4.4.1 and 4.4.2, Mike, Kevin, Julia, and Adam understood GenAI tools as *"word generating machines"* (Adam). Mike explained, *"I think it works on the word prediction or language prediction model... I would probably call it an advanced parrot who can understand language. It understands the patterns and the technicalities behind the language, how the language works."* Kevin also believed that ChatGPT could predict the next likely word given an input sequence of words. However, he held a more skeptic view of GenAI:

> *"When I throw a sentence into it… it spits out connected words that kind of go in the direction of the prompt. There is clearly no overarching intelligence in there. It just comes up with words that string together and they kind of sound vaguely intelligent based off what you're trying to ask it."*

Unsurprisingly, participants who shared this mental model had programming backgrounds themselves or partners who worked in the technology sector.

4.4.4 *Stores and Reuses User Prompts.*

Some participants thought that GenAI tools gathered information from users' conversation histories. Sara explained,

> *"It will learn from what other people put in, I think. If I ask it to write me an email... maybe it is able to grab from someone else who asked the same question or someone who wrote a similar email. It probably stores stuff that we all do, anonymously in some way, in order for it to learn what people want and what people like."*

Important to note is that participants here did not refer to the mechanism of user feedback in GenAI tools; in fact, many participants did not know about upvote/downvote buttons because those were unlabeled and inaccessible (see Section 4.1). In this case, participants understood users' conversation histories to be contributing to other sources of information used by GenAI (e.g., the large database mentioned in Section 4.4.2).

4.4.5 *"More In-depth AI".*

Participants conceptualized how GenAI worked by comparing them with other applications they had used for accomplishing similar tasks. They compared text-based GenAI like ChatGPT with voice assistants (e.g., Alexa, Siri, Google Assistant) and Be My AI with other image description apps (e.g., Seeing AI, Tap Tap See). In both cases, they considered GenAI to be *"more in-depth AI"* (Ivan) for their ability to provide comprehensive information in a more conversational way. In fact, when contrasted with GenAI, several participants critiqued voice assistants and older image description apps for being *"not too intelligent"* (Lily) and questioned whether those apps can be truly characterized as AI.

4.4.6 *Partner, Friend, Mentor, Secretary...*

Participants used several anthropomorphized [21, 44] metaphors to describe the ways in which they conceived of GenAI tools, such as *"your notetaker slash secretary you never had"* (Gina), *"a librarian working for you"* (Frank), *"a friend who's your personal proofreader… and who can be honest to help clarify what you're really trying to say"* (Portia), and *"a writing assistant, editor… best friend, and mentor kind of thing that never judges and is never in a bad mood"* (Julia). To Henry, anthropomorphization extended beyond GenAI capabilities to their idiosyncratic behaviors. He said,

> *"I just find it quite amusing that sometimes I'll have good days with it, sometimes I have bad days with it and I'm not sure what's causing that. That's very human—humans have off days and good days, but I can't imagine why that would happen with AI."*

Adam used anthromorphized metaphors to describe how he had developed certain levels of trust and comfort with GenAI tools over time by *"treating it like a partner in getting things done... At the beginning, you're getting used to each other's quirks... By the time you've been working together for a while, you know how each other work, what you can trust and what you can't."*

4.4.7 *"Still a Computer, Not a Human".*

Despite some anthropomorphization, upon deeper reflection, Ethan, Adam, Daisy, and Mike agreed that GenAI tools are *"just a computer. It'll never be as good as a human."* This mental model was informed by situations where GenAI tools faltered at solving *"mathematical and logical problems"* (Mike) and issues that require *"nuanced reasoning"* (Daisy), such as determining words containing certain letters for an anagram game or solving complicated coding problems. Adam explained, *"It doesn't think like a regular developer does… So, bugs [in codes] that might seem obvious to a user or developer… through the amount of experience we've had coding, might not seem obvious to it."* Likewise, participants believed GenAI tools could not perform well at addressing requests that require *"creativity"* (Daisy, Adam, Noah) or expressing *"subjective opinions"* (Lily, Noah, Adam), such as writing novels or presenting arguments on whether mountains or oceans are better for vacations. Adam shared that the reason GenAI tools are not creative lies in the way they are fundamentally constructed.

> *"This is not an idea generating machine… If I wanted it to write about the quests that the characters [in a story] undertook, it's going to regurgitate very common themes from classic lit… It's not going to come out with a new one out of whole cloth that's gonna turn anyone into a bestseller author… It does not have a human spark of creativity, doesn't think outside the box."*

Collectively, these mental models reveal the ways in which our participants conceptualized how GenAI tools retrieve and store information and generate responses, drawing comparisons with other AI systems and even humans. These mental models were rooted in their firsthand experience with GenAI as blind screen reader users, as are their perceptions about harms and biases of GenAI, which we elaborate on in the next section.

4.5 Reflecting on Biases and Harms of Generative AI

Aligning with prior work [25, 29, 60], our analysis foregrounds how blind individuals think through biases and harms of GenAI while using these tools for content creation. Participants recounted several instances where GenAI tools could not *"handle nuanced concepts"* (Daisy) about disability and produced ableist and ageist content. During our sessions, Noah and Gina prompted ChatGPT to create a story involving a blind person traveling to a new country. In the stories, ChatGPT described the blind person as 'courageous' and 'resilient.' Noting this characterization of blind people as *"stereotypical,"* Noah elaborated on why certain linguistic choices by GenAI might be subtly inappropriate: *"It's not a major deal breaker... But I can hear some of my blind friends [say] we're not all courageous."* Daisy recalled asking ChatGPT to generate a story about a *"disabled person going on an adventure. It couldn't do that. It was like, 'Here's a family of adventurers, but then they had a disabled person and that person stayed home.'"* Similarly, while brainstorming ideas for messages to include in a birthday card for her mother, Carla prompted ChatGPT *"to give a compliment about being old… but it made a couple of negative comments about being old."*

GenAI tools exibited ableism not only in the generated content but also in their interaction with users. Ethan, Bella, and Carla encountered situations where they were looking for blindness-related information on ChatGPT and Copilot and the tools expressed grief for their disability, saying "*I'm sorry, you're blind.*" This frustrated Bella: *"Seriously, can we move on? I don't really need the AI thing apologizing to me because I'm blind."* Participants tried to correct GenAI by explaining in the chat why those reactions were inappropriate, because they believed that chat history stored by GenAI tools would be reused for future improvements (see Section 4.4.4). Carla elaborated, *"I'll start with trying to give a correction, like you shouldn't tell blind people that you feel sorry for them... in the hopes that that would be incorporated in the future."* Thus, blind participants expended considerable effort in providing feedback to mitigate ableist GenAI responses, which exemplifies the significant advocacy labor disabled individuals must perform to voice their needs and reduce equity gaps reified by technologies [80].

Echoing findings in prior work [25], our participants hypothesized biased dataset as a key reason behind ableist and inappropriate GenAI responses. Kevin explained, *"The dataset is probably mostly [nondisabled] people writing about us rather than people in the disabled community."* Carla thought that GenAI responses might be driven by prevalent *"misunderstandings about particular disabilities"* and would further reinforce those misconceptions, such as blind people desiring to *"feel [someone's] face... to help them visualize"* how others look like. Given the impact of biased datasets on GenAI output, participants felt that *"feeding the model large amounts of data written from the disability perspective would be good"* (Kevin).

Besides issues related to disability and ableism, our participants were cognizant of GenAI showing biased portrayals of other aspects of identity like race and gender. In one instance, Carla noticed that Be My AI were *"assuming short hair meant a boy as opposed to a girl."* Similarly, Kevin proactively edited ChatGPT responses when it misgendered somebody or described one's disability in a way not preferred by them. As another example of biases against underrepresented populations, during our session, Adam asked ChatGPT to formulate sentences in Maori language and found two out of ten resultant Maori sentences to be grammatically incorrect. He speculated the lack of representative dataset as the reason behind this: *"Maori is a very low resource language. There aren't a lot of people that put it on the web. It's very underrepresented in the datasets. So, it's not gonna know as much about it as it will know English."* Daisy critiqued issues around AI fairness and bias more broadly:

> *"We call them artificial intelligence, but they are ultimately based on humans and humans have internal biases. And the disabled community and the community of minorities face bias every day. And so, these artificial intelligence models that are being built, when they are searching the internet, their sources are going to be impacted by bias... racism, ableism, and so on."*

Reflecting on GenAI's other negative impacts, participants were concerned about the proliferation of mis- and disinformation through GenAI. Sara, Henry, Mike, and Ethan shared that the *"biggest fear"* of the GenAI boom was the rise in propaganda, deep fake videos and images impersonating people without their consent, and the use of voice cloning AI to scam others. Henry said, *"Usually I'm very good at kind of detecting scams but [voice cloning] is one scam I don't think I will be able to detect."* Our participants maintained extra caution before using GenAI responses in content they would publicly share, because they did not want to *"create more fake news in the world"* (Kevin). Julia echoed this sentiment, saying: *"If I'm alt tagging [an image using Be My AI] for purposes of others being able to access it with a screen reader, I will check with someone sighted… to make sure that everything is described correctly because I don't want to give misinformation."* Portia, Theo, Henry, and Ethan were apprehensive about privacy issues due to a limited understanding about whether GenAI tools stored their information, for what purposes and how long, and how that information would be used [98]. The ability to retrieve previous chat history fueled their concerns.

While reasoning through these promises and perils of GenAI on individual, interpersonal, and social level, our participants expressed willingness to embrace the growth in GenAI. They were cognizant of and concerned about potential harms of GenAI; however, they did not want GenAI's progress to be stifled, given its positive impacts on enhancing and scaling accessibility [30, 37, 86]. Sara said, *"The worst that could happen is it would all just go away and people would stop developing it, and it would be sort of like something that we had for a minute, and it was great and then it just sort of fizzled out."* Overall, these perspectives from blind participants have implications for future efforts to reconcile the positive and negative effects of GenAI tools, which we will revisit in the Discussion.

5 DISCUSSION

We have presented one of the first detailed accounts of the practices and mental models of generative AI among blind people. Below we synthesize our findings to rethink access and information verification in GenAI, unpack erroneous mental models of GenAI among blind individuals, and reflect on harms and benefits of GenAI through an accessibility-centric lens.

5.1 Rethinking Access and Information Verification in Generative AI

Building accessible technology requires critically considering not only whether disabled people can access it on a basic level but also the extent to which they can leverage the full benefits of these technologies. As our analysis demonstrates, the inherent accessibility of text-based interaction enables basic levels of nonvisual access in current GenAI tools; however, these tools still leave a lot to be desired for blind users due to the lack of accessible keyboard navigation, unlabeled buttons, and poor UI design. For example, the suboptimal and inaccessible UI of some GenAI tools required our participants to piece together a cumbersome workflow involving multiple GenAI and writing applications just to read and edit AI-generated responses. For blind users, such issues greatly diminish the efficiency gained from using GenAI, which has been positioned as one of the biggest benefits of using it for content creation [45, 84].

On top of this, inaccuracies and hallucinations in GenAI [11, 83] add additional layers of complexity for blind users. The possibility of inaccurate responses requires users to decide whether and to what extent they need to verify the generated responses, and we observed four factors that shaped this decision: *context of use*, *stakes*, *verifiability*, and *believability*. While all users—blind and sighted alike—need to grapple with these factors to counteract GenAI inaccuracies, we argue that blind individuals' use cases require unique considerations. Moreover, accessibility issues of GenAI tools can severely limit which of these factors they can prioritize when deciding whether to verify a response, adversely affecting their likelihood of avoiding misinformation [81, 100].

First, with regards to the *context of use*, we observed that blind users are more likely to verify GenAI responses that would be used in medical, health, education, financial, professional, or public contexts compared to personal use cases. Second, even within the same context, blind users consider whether the *stakes* are high enough to justify verification. For instance, when generating image descriptions for food labels during cooking, the verification stakes are higher for information about the presence of allergens or the expiry date of a product than for other information with higher tolerance for error, such as heating time or how much seasoning to add. Similarly, in the context of sharing images on social media, the AI-generated alt text of a personal photo shared for fun purposes has lower stakes for accuracy than the one for an infographic containing important health-related information. Third, users consider how easily GenAI responses can be verified to judge whether verification is worth the effort (i.e., *verifiability*) [57]. When producing image description using GenAI (e.g., By My AI), sighted users can readily determine mismatches between visual content and the generated description, whereas blind users need to seek sighted help, significantly reducing verifiability for blind users in situations where sighted help is unavailable or inconvenient. Even for textual responses given by GenAI, efforts needed to verify (e.g., through another search engine) can be higher for blind individuals due to the inaccessibility and usability issues associated with copying and pasting responses across different platforms [59, 89]. Finally, blind users rely on perceived *believability* [70, 91] of GenAI responses for deciding whether or not to verify the information presented. Our participants shared that they often forego fact-checking if GenAI responses do not *"seem fishy"* or *"unexpected."* As prior work found, blind users tend to unduly trust auto-generated image descriptions [61]. GenAI tools amplify this issue since their responses are relatively richer, more comprehensive, and detailed, which increases the believability of these responses compared to other image recognition applications (e.g., Be My AI vs Seeing AI).

Thus, the richness of the GenAI responses becomes a double-edged sword, which on one hand significantly improves access to visual information while also making it more likely for people to trust inaccurate information. The issues of stake and verifiability have also been discussed by Glazko et al. [30]. Our analysis reveals believability as yet another key factor that problematizes the verification decision-making process for blind users, especially those who may not be knowledgeable of the idiosyncrasies of GenAI and are less likely to be skeptical about generated responses. These issues are not confined to visual information only but apply more broadly to all GenAI tools. For instance, references to source websites, as tools like Microsoft Copilot and Perplexity include in their responses, reinforce their perceived believability among our participants. While citing references may seem to help users understand provenance i.e., source of the information presented, researchers have found that GenAI-cited references are often inaccurate or do not substantiate the associated statements [57, 63].

It is also important to highlight that the four factors stated above are not siloed, rather they are often competing or at tension with each other. For example, when stakes are higher, a user would be willing to verify a response despite high believability. Conversely, even if a generated image description has slightly low believability, a blind user might be willing to forgo verification because of low verifiability, such as due to the unavailability of sighted help.

Given these issues and tensions, we argue that researchers and developers must work toward reducing frictions that minimize the benefits of using GenAI for blind users. In addition to enforcing established accessibility principles within individual GenAI tools, we suggest that further attention be given toward tailoring the reading and editing experiences within GenAI tools for nonvisual access so that the effort needed from blind individuals to review or edit generated content (which currently requires switching back and forth between multiple apps) does not diminish the benefits they receive from using GenAI. More importantly, we feel there is an acute need for seamless ways to verify information in GenAI tools so that blind users do not need to expend significantly extra effort for verification. Developers may consider integrating strategies blind users already adopt to further streamline their verification workflow. For example, GenAI tools may test its response consistency across repeated tries in the same tool or in multiple tools and summarize these inconsistencies for blind users which may encourage constructive skepticism among users [18] towards otherwise believable GenAI responses.

5.2 Unpacking (Erroneous) Mental Models of Generative AI among Blind People

Our analysis reveals that blind users often develop flawed, incomplete, or oversimplified mental models of GenAI tools [55, 71, 95]. Several of these mental models align with that of sighted users, as found in recent work [98]. For instance, our participants' mental models of GenAI chatbots, even ChatGPT 3.5, as collecting information from the internet through a search engine (Section 4.4.1) or employing keyword-based search on a massive database (Section 4.4.2) match sighted users' understanding of ChatGPT as a "super searcher." Similarly, those with more technical know-how (both blind and sighted [98]) understood GenAI chatbots as *"advanced parrots"* or stochastic *"word generating machines."* Our analysis also reveals potentially severe implications of erroneous mental models [55]. For example, most blind participants had a misconception that GenAI tools were good at answering factual questions [57, 63, 94], informed by the belief that these tools (even the ones without actual online access) always pulled up the most relevant and accurate information available on the internet.

Importantly, blind users' mental models are also shaped by their experience with blindness, the assistive technologies they use, and the level of accessibility in GenAI tools, which may differ from the way sighted users form mental models. For example, sighted users understood ChatGPT to be incorporating quality-related feedback provided through the upvote/downvote buttons [98]. In contrast, many blind participants were unaware of the upvote/downvote buttons because those were unlabeled and inaccessible. They instead provided feedback by typing follow-up prompts in the chat (e.g., asking ChatGPT to not express grief for someone's blindness), and because they could go back to their conversation history later, they assumed that ChatGPT would utilize these follow-up corrections in future conversations to improve responses.

This juxtaposition between mental models of blind and sighted users from a user experience perspective shows how accessibility oversights as simple as an unlabeled button may lead to divergent and erroneous mental models among blind users. We argue that any measures taken to improve transparency and trust of users for GenAI tools [55, 79] (e.g., UI redesign to minimize deceptive or dark patterns [98]) must be examined from the perspective of nonvisual access. This is just one example of how flawed mental models can be shaped by inaccessible design even on a simple, text-based chat interface; however, as GenAI interfaces continue to evolve and integrate more complex, multimodal features [64], supporting blind users in forming or shifting to accurate mental models will remain a critical challenge in AI and accessibility.

Furthermore, our participants' anthropomorphized perception of GenAI as *"best friend"* or *"partner in getting things done"* may have contributed to their heightened (and often misplaced) trust on these tools. However, the ways in which anthropomorphized descriptions of AI influence the public's trust and reliance on these tools are complicated [44]. As such, further research is required to uncover the nuances of anthropomorphization, trust, and mental models of GenAI among blind people and how these aspects are shaped by the publicity and media representation of GenAI capabilities [55] and their impacts on accessibility.

5.3 Reconsidering Harms and Benefits of Generative AI through the Lens of Access

Our analysis joins that of others who call attention to harmful disability representations in GenAI, both in text produced by LLM chatbots [25, 29] and images generated with text-to-image models [60]. We bring out empirically-driven insights from blind participants' everyday experiences, reconfirming the prevalence of ableist and ageist biases in GenAI. For instance, our participants encountered stereotypical characterization of blind people (e.g., courageous, resilient) that bordered on 'inspiration porn' [34], i.e., languages that objectify disabled people as being inspirational for the gratification of non-disabled people. Important to note is that biases exist in not only the content produced by GenAI but also the ways these tools interact with users. Participants found that often questions about ideas for performing a task as a blind person are met with ChatGPT expressing pity and grief about their blindness. Thus, in addition to highlighting the need for more representative datasets [25, 30], we argue that researchers and developers need to critically examine and update GenAI tools' default response behavior such that it is not codified to reinforce ableist narratives about disability.

Furthermore, we call for a nuanced approach in addressing biases and harms around GenAI such that measures taken to alleviate harms do not disregard the accessibility support disabled people receive from GenAI tools [37]. Our participants emphasized how GenAI helped them address critical needs, such as spellchecking and formatting support while writing and coding as well as getting detailed visual information—areas where existing systems are inaccessible and extremely challenging to navigate for screen reader users [19, 53, 67]. Thus, while the use of GenAI in academic writing brings forth legitimate concerns around plagiarism [17, 69], implementing extreme measures (such as outright bans on GenAI use at schools) to counteract plagiarism may deprive blind students of the accessibility benefits of GenAI such as automated proofreading. Another recent example of this is the public outcry on social media when Be My AI stopped describing images with people's faces [23].[2] This exemplifies the tensions around the ways in which addressing harms in one dimension (e.g., privacy violation due to facial recognition in images [82]) may reify harms in another dimension (e.g., revoking the accessibility benefits for blind users, which might be construed as a quality-of-service harm [82]). Finding an ethical and productive way forward to combat GenAI-related harms without minimizing the progress in accessibility is a tough challenge that does not have a clear-cut solution but one that researchers, designers, and policymakers in AI, fairness, and accessibility must approach thoughtfully and carefully together.

5.4 Limitations and Future Work

An important limitation of our study is that most of our participants were intermediate users of GenAI, although we had a few experts and one beginner. Future studies can specifically focus on the experiences of blind people who are beginner or expert users to uncover the similarities and differences in their usage patterns and mental models. Additionally, we purposefully kept our focus broad to reveal the general GenAI usage patterns among blind users across diverse contexts including information retrieval, coding, copywriting, creative writing, and more. Future work can extend our findings by investigating nuanced practices and accessibility within specific use cases. Finally, our analysis only focused on blind users' experience with text-based and image-description GenAI tools. Further research is needed to explore how blind people interact with multimodal GenAI tools, e.g., image and music generation to develop a holistic understanding of GenAI accessibility.

6 CONCLUSION

Through our inquiry into Generative AI (GenAI) usage of blind individuals, we uncover in-depth empirical understandings of the diverse ways in which blind people use GenAI to streamline their content creation and information retrieval workflows, often working around various accessibility and usability issues in GenAI interfaces. Through this, we shed light on the complex cost-benefit analysis

[2]A later update reverted this change and as of the writing of this paper, Be My AI was describing images with faces again [77].

blind users perform to navigate the inaccuracies and idiosyncrasies of GenAI tools while managing the effort for information verification. Additionally, we reveal blind individuals' mental models of GenAI systems which both align with and differ from that of sighted users but are often erroneous and oversimplified nonetheless. We argue that to enable equitable opportunities for blind individuals to leverage the benefits of GenAI, we must revisit the design and policy discussions around supporting users in building accurate mental models, verifying information accuracy, and combating biases and harms of GenAI through the lens of nonvisual access.

7 ACKNOWLEDGMENT

We thank our participants for their contributions to this study. We are grateful to Abir Saha for his feedback on multiple iterations of this draft. We also thank Cynthia Bennett for a thought-provoking conversation about GenAI harms.

REFERENCES

[1] Ali Abdolrahmani, Kevin M. Storer, Antony Rishin Mukkath Roy, Ravi Kuber, and Stacy M. Branham. 2020. Blind Leading the Sighted: Drawing Design Insights from Blind Users towards More Productivity-oriented Voice Interfaces. *ACM Transactions on Accessible Computing (TACCESS)* 12, 4, Article 18 (jan 2020), 35 pages. https://doi.org/10.1145/3368426

[2] Gus Alexiou. 2023. Envision Adds ChatGPT AI Sight Assistance To Its Smart Glasses For The Blind. Retrieved 23 April, 2024 from https://www.forbes.com/sites/gusalexiou/2023/04/30/envision-adds-chatgpt-ai-sight-assistance-to-its-smart-glasses-for-the-blind/?sh=d443c835446f.

[3] Safinah Ali, Prerna Ravi, Randi Williams, Daniella DiPaola, and Cynthia Breazeal. 2024. Constructing Dreams Using Generative AI. *Proceedings of the AAAI Conference on Artificial Intelligence* 38, 21 (Mar. 2024), 23268–23275. https://doi.org/10.1609/aaai.v38i21.30374

[4] Matin Amoozadeh, David Daniels, Daye Nam, Aayush Kumar, Stella Chen, Michael Hilton, Sruti Srinivasa Ragavan, and Mohammad Amin Alipour. 2024. Trust in Generative AI among Students: An exploratory study. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1* (Portland, OR, USA) *(SIGCSE 2024)*. ACM, 67–73. https://doi.org/10.1145/3626252.3630842

[5] Anthropic. 2024. Claude: Conversational AI by Anthropic. Retrieved 23 April, 2024 from https://www.anthropic.com/claude.

[6] Gagan Bansal, Besmira Nushi, Ece Kamar, Walter S Lasecki, Daniel S Weld, and Eric Horvitz. 2019. Beyond Accuracy: The Role of Mental Models in Human-AI Team Performance. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 7. 2–11. https://doi.org/10.1609/hcomp.v7i1.5285

[7] Edward C Bell and Natalia M Mino. 2015. Employment Outcomes for Blind and Visually Impaired Adults. *Journal of Blindness Innovation and Research* 5, 2 (2015). http://dx.doi.org/10.5241/5-85

[8] Nathan Bos, Kimberly Glasgow, John Gersh, Isaiah Harbison, and Celeste Lyn Paul. 2019. Mental models of AI-based systems: User predictions and explanations of image classification results. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 63. 184–188. https://doi.org/10.1177/1071181319631392

[9] Stephen Brade, Bryan Wang, Mauricio Sousa, Sageev Oore, and Tovi Grossman. 2023. Promptify: Text-to-Image Generation through Interactive Prompt Exploration with Large Language Models. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (San Francisco, CA, USA) *(UIST '23)*. ACM, Article 96, 14 pages. https://doi.org/10.1145/3586183.3606725

[10] Virginia Braun and Victoria Clarke. 2021. *Thematic analysis: a practical guide.* SAGE Publications Ltd. https://us.sagepub.com/en-us/nam/thematic-analysis/book248481

[11] Ritvik Budhiraja, Ishika Joshi, Jagat Sesh Challa, Harshal D. Akolekar, and Dhruv Kumar. 2024. "It's not like Jarvis, but it's pretty close!" - Examining ChatGPT's Usage among Undergraduate Students in Computer Science. In *Proceedings of the 26th Australasian Computing Education Conference* (Sydney, NSW, Australia) *(ACE '24)*. ACM, 124–133. https://doi.org/10.1145/3636243.3636257

[12] Raluca Budiu. 2024. Generative AI Diary. Retrieved 23 April, 2024 from https://www.nngroup.com/articles/generative-ai-diary/.

[13] Nicholas Carlini, Florian Tramèr, Eric Wallace, Matthew Jagielski, Ariel Herbert-Voss, Katherine Lee, Adam Roberts, Tom Brown, Dawn Song, Úlfar Erlingsson, Alina Oprea, and Colin Raffel. 2021. Extracting Training Data from Large Language Models. In *30th USENIX Security Symposium (USENIX Security 21)*. USENIX Association, 2633–2650. https://www.usenix.org/conference/usenixsecurity21/presentation/carlini-extracting

[14] Tiffany Chen, Cassandra Lee, Jessica R Mindel, Neska Elhaouij, and Rosalind Picard. 2023. Closer Worlds: Using Generative AI to Facilitate Intimate Conversations. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI EA '23)*. ACM, Article 68, 15 pages. https://doi.org/10.1145/3544549.3585651

[15] Yuyan Chen, Qiang Fu, Yichen Yuan, Zhihao Wen, Ge Fan, Dayiheng Liu, Dongmei Zhang, Zhixu Li, and Yanghua Xiao. 2023. Hallucination Detection: Robustly Discerning Reliable Answers in Large Language Models. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management* (Birmingham, United Kingdom) *(CIKM '23)*. ACM, 245–255. https://doi.org/10.1145/3583780.3614905

[16] Ruijia Cheng, Ruotong Wang, Thomas Zimmermann, and Denae Ford. 2024. "It would work for me too": How Online Communities Shape Software Developers' Trust in AI-Powered Code Generation Tools. *ACM Transactions on Interactive Intelligent Systems (TIIS)* 14, 2, Article 11 (may 2024), 39 pages. https://doi.org/10.1145/3651990

[17] Debby R. E. Cotton, Peter A. Cotton, and J. Reuben Shipway. 2024. Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International* 61, 2 (2024), 228–239. https://doi.org/10.1080/14703297.2023.2190148

[18] Maitraye Das, Alexander J. Fiannaca, Meredith Ringel Morris, Shaun K. Kane, and Cynthia L. Bennett. 2024. From Provenance to Aberrations: Image Creator and Screen Reader User Perspectives on Alt Text for AI-Generated Images. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. 21 pages. https://doi.org/10.1145/3613904.3642325

[19] Maitraye Das, Darren Gergle, and Anne Marie Piper. 2019. "It doesn't win you friends": Understanding Accessibility in Collaborative Writing for People with Vision Impairments. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW, Article 191 (nov 2019), 26 pages. https://doi.org/10.1145/3359293

[20] Michael Ann DeVito, Jeremy Birnholtz, Jeffery T. Hancock, Megan French, and Sunny Liu. 2018. How People Form Folk Theories of Social Media Feeds and What it Means for How We Study Self-Presentation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. ACM, 1–12. https://doi.org/10.1145/3173574.3173694

[21] Nicholas Epley, Adam Waytz, and John T Cacioppo. 2007. On seeing human: a three-factor theory of anthropomorphism. *Psychological review* 114, 4 (2007), 864. https://doi.org/10.1037/0033-295X.114.4.864

[22] 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* (San Jose, California, USA) *(CHI '16)*. ACM, 2371–2382. https://doi.org/10.1145/2858036.2858494

[23] Be My Eyes. 2023. Image to Text AI and Blurred Faces: What's Going On? Retrieved April 23, 2024 from https://www.bemyeyes.com/blog/image-to-text-ai-and-blurred-faces-whats-going-on.

[24] Stefan Feuerriegel, Jochen Hartmann, Christian Janiesch, and Patrick Zschech. 2024. Generative AI. *Business & Information Systems Engineering* 66, 1 (2024), 111–126. https://link.springer.com/article/10.1007/s12599-023-00834-7

[25] 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* (Chicago, IL, USA) *(FAccT '23)*. ACM, 205–216. https://doi.org/10.1145/3593013.3593989

[26] Katy Ilonka Gero, Zahra Ashktorab, Casey Dugan, Qian Pan, James Johnson, Werner Geyer, Maria Ruiz, Sarah Miller, David R. Millen, Murray Campbell, Sadhana Kumaravel, and Wei Zhang. 2020. Mental Models of AI Agents in a Cooperative Game Setting. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. ACM, 1–12. https://doi.org/10.1145/3313831.3376316

[27] Katy Ilonka Gero, Tao Long, and Lydia B Chilton. 2023. Social Dynamics of AI Support in Creative Writing. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. ACM, Article 245, 15 pages. https://doi.org/10.1145/3544548.3580782

[28] Sarah Gibbons, Tarun Mugunthan, and Jakob Nielsen. 2023. Accordion Editing and Apple Picking: Early Generative-AI User Behaviors. Retrieved 23 April, 2024 from https://www.nngroup.com/articles/accordion-editing-apple-picking/.

[29] Kate Glazko, Yusuf Mohammed, Ben Kosa, Venkatesh Potluri, and Jennifer Mankoff. 2024. Identifying and Improving Disability Bias in GPT-Based Resume Screening. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency* (Rio de Janeiro, Brazil) *(FAccT '24)*. ACM, 687–700. https://doi.org/10.1145/3630106.3658933

[30] 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* (New York, NY, USA) *(ASSETS '23)*. ACM, Article 99, 8 pages. https://doi.org/10.1145/3597638.3614548

[31] Ricardo E Gonzalez Penuela, Jazmin Collins, Cynthia Bennett, and Shiri Azenkot. 2024. Investigating Use Cases of AI-Powered Scene Description Applications for Blind and Low Vision People. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. ACM, Article 901, 21 pages. https://doi.org/10.1145/3613904.3642211

[32] Steven M. Goodman, Erin Buehler, Patrick Clary, Andy Coenen, Aaron Donsbach, Tiffanie N. Horne, Michal Lahav, Robert MacDonald, Rain Breaw Michaels, Ajit Narayanan, Mahima Pushkarna, Joel Riley, Alex Santana, Lei Shi, Rachel Sweeney, Phil Weaver, Ann Yuan, and Meredith Ringel Morris. 2022. LaMPost: Design and Evaluation of an AI-assisted Email Writing Prototype for Adults with Dyslexia. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility* (Athens, Greece) *(ASSETS '22)*. ACM, Article 24, 18 pages. https://doi.org/10.1145/3517428.3544819

[33] Google. 2024. Gemini. Retrieved 12 April, 2024 from https://gemini.google.com/app.

[34] Jan Grue. 2016. The problem with inspiration porn: a tentative definition and a provisional critique. *Disability & Society* 31, 6 (2016), 838–849. https://doi.org/10.1080/09687599.2016.1205473

[35] 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 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, Los Alamitos, CA, USA, 3608–3617. https://doi.org/10.1109/CVPR.2018.00380

[36] Orit Hazzan and Yael Erez. 2024. Generative AI in Computer Science Education. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 2* (Portland, OR, USA) *(SIGCSE 2024)*. ACM, 1899. https://doi.org/10.1145/3626253.3633409

[37] Amanda Heidt. 2024. 'Without these tools, I'd be lost': how generative AI aids in accessibility. Retrieved 23 April, 2024 from https://www.nature.com/articles/d41586-024-01003-w.

[38] Laurie Henneborn. 2023. Designing Generative AI to Work for People with Disabilities. Retrieved 23 April, 2024 from https://hbr.org/2023/08/designing-generative-ai-to-work-for-people-with-disabilities.

[39] 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* (Athens, Greece) *(ASSETS '22)*. ACM, Article 14, 18 pages. https://doi.org/10.1145/3517428.3544824

[40] Aike C. Horstmann, Clara Strathmann, Lea Lambrich, and Nicole C. Krämer. 2023. Alexa, What's Inside of You: A Qualitative Study to Explore Users' Mental Models of Intelligent Voice Assistants. In *Proceedings of the 23rd ACM International Conference on Intelligent Virtual Agents* (Würzburg, Germany) *(IVA '23)*. ACM, Article 5, 10 pages. https://doi.org/10.1145/3570945.3607335

[41] Irene Hou, Sophia Mettille, Owen Man, Zhuo Li, Cynthia Zastudil, and Stephen MacNeil. 2024. The Effects of Generative AI on Computing Students' Help-Seeking Preferences. In *Proceedings of the 26th Australasian Computing Education Conference* (Sydney, NSW, Australia) *(ACE '24)*. ACM, 39–48. https://doi.org/10.1145/3636243.3636248

[42] Yue Huang, Borke Obada-Obieh, and Konstantin (Kosta) Beznosov. 2020. Amazon vs. My Brother: How Users of Shared Smart Speakers Perceive and Cope with Privacy Risks. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. ACM, 1–13. https://doi.org/10.1145/3313831.3376529

[43] Mina Huh, Yi-Hao Peng, and Amy Pavel. 2023. GenAssist: Making Image Generation Accessible. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (San Francisco, CA, USA) *(UIST '23)*. ACM, Article 38, 17 pages. https://doi.org/10.1145/3586183.3606735

[44] Nanna Inie, Stefania Druga, Peter Zukerman, and Emily M. Bender. 2024. From "AI" to Probabilistic Automation: How Does Anthropomorphization of Technical Systems Descriptions Influence Trust?. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency* (Rio de Janeiro, Brazil) *(FAccT '24)*. ACM, 2322–2347. https://doi.org/10.1145/3630106.3659040

[45] Nanna Inie, Jeanette Falk, and Steve Tanimoto. 2023. Designing Participatory AI: Creative Professionals' Worries and Expectations about Generative AI. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI EA '23)*. ACM, Article 82, 8 pages. https://doi.org/10.1145/3544549.3585657

[46] JiWoong Jang, Sanika Moharana, Patrick Carrington, and Andrew Begel. 2024. "It's the only thing I can trust": Envisioning Large Language Model Use by Autistic Workers for Communication Assistance. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. ACM, Article 77, 18 pages. https://doi.org/10.1145/3613904.3642894

[47] 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* (Hamburg, Germany) *(CHI '23)*. ACM, Article 827, 18 pages. https://doi.org/10.1145/3544548.3581337

[48] Ruogu Kang, Laura Dabbish, Nathaniel Fruchter, and Sara Kiesler. 2015. "My data just goes everywhere": User mental models of the internet and implications for privacy and security. In *Proceedings of the Eleventh USENIX Conference on Usable Privacy and Security* (Ottawa, Canada) *(SOUPS '15)*. USENIX Association, USA, 39–52. https://www.usenix.org/conference/soups2015/proceedings/presentation/kang

[49] Majeed Kazemitabaar, Justin Chow, Carl Ka To Ma, Barbara J. Ericson, David Weintrop, and Tovi Grossman. 2023. Studying the effect of AI Code Generators on Supporting Novice Learners in Introductory Programming. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. ACM, Article 455, 23 pages. https://doi.org/10.1145/3544548.3580919

[50] David E Kieras and Susan Bovair. 1984. The role of a mental model in learning to operate a device. *Cognitive science* 8, 3 (1984), 255–273. https://doi.org/10.1207/s15516709cog0803_3

[51] Tanya Kraljic and Michal Lahav. 2024. From Prompt Engineering to Collaborating: A Human-Centered Approach to AI Interfaces. *Interactions* 31, 3 (may 2024), 30–35. https://doi.org/10.1145/3652622

[52] Sri H. Kurniawan, Alistair G. Sutcliffe, Paul L. Blenkhorn, and Jae Eun Shin. 2003. Mental Models of Blind Users in the Windows Environment. *International Journal of Rehabilitation Research* 26, 2 (June 2003), 145–147. https://doi.org/10.1097/00004356-200306000-00011

[53] Nimatul Lailiyah, Iswahyuni Iswahyuni, and Alies Poetri Lintangsari. 2020. English Spelling Errors Committed by the Blind. *Indonesian Journal of Disability Studies* 7, 1 (May 2020), 66–71. https://doi.org/10.21776/ub.ijds.2019.007.01.7

[54] Jenny T. Liang, Chenyang Yang, and Brad A. Myers. 2024. A Large-Scale Survey on the Usability of AI Programming Assistants: Successes and Challenges. In *Proceedings of the IEEE/ACM 46th International Conference on Software Engineering* (Lisbon, Portugal) *(ICSE '24)*. ACM, Article 52, 13 pages. https://doi.org/10.1145/3597503.3608128

[55] Q. Vera Liao and Jennifer Wortman Vaughan. 2024. AI Transparency in the Age of LLMs: A Human-Centered Research Roadmap. *Harvard Data Science Review* Special Issue 5 (feb 29 2024). https://hdsr.mitpress.mit.edu/pub/aelql9qy

[56] Lauren Lin and Duri Long. 2023. Generative AI Futures: A Speculative Design Exploration. In *Proceedings of the 15th Conference on Creativity and Cognition* (Virtual Event, USA) *(C&C '23)*. ACM, 380–383. https://doi.org/10.1145/3591196.3596616

[57] Nelson F. Liu, Tianyi Zhang, and Percy Liang. 2023. Evaluating Verifiability in Generative Search Engines. In *Conference on Empirical Methods in Natural Language Processing (EMNLP '23)*. https://doi.org/10.48550/arXiv.2304.09848

[58] Duri Long and Brian Magerko. 2020. What is AI Literacy? Competencies and Design Considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. ACM, 1–16. https://doi.org/10.1145/3313831.3376727

[59] Darren Lunn, Simon Harper, and Sean Bechhofer. 2011. Identifying Behavioral Strategies of Visually Impaired Users to Improve Access to Web Content. *ACM ACM Transactions on Accessible Computing (TACCESS)* 3, 4, Article 13 (April 2011), 35 pages. https://doi.org/10.1145/1952388.1952390

[60] Avery Mack, Rida Qadri, Remi Denton, Shaun K Kane, and Cynthia L Bennett. 2024. "They only care to show us the wheelchair": disability representation in text-to-image AI models. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. 21 pages. https://doi.org/10.1145/3613904.3642166

[61] 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* (Denver, Colorado, USA) *(CHI '17)*. ACM, 5988–5999. https://doi.org/10.1145/3025453.3025814

[62] Kamil Malinka, Martin Peresíni, Anton Firc, Ondrej Hujnák, and Filip Janus. 2023. On the Educational Impact of ChatGPT: Is Artificial Intelligence Ready to Obtain a University Degree?. In *Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education V. 1* (Turku, Finland) *(ITiCSE 2023)*. ACM, 47–53. https://doi.org/10.1145/3587102.3588827

[63] Shahan Ali Memon and Jevin D. West. 2024. Search engines post-ChatGPT: How generative artificial intelligence could make search less reliable. Retrieved April 27, 2024 from https://www.cip.uw.edu/2024/02/18/search-engines-chatgpt-generative-artificial-intelligence-less-reliable/

[64] Meta AI. 2023. Multimodal Generative AI Systems. Retreived 23 April, 2024 from https://ai.meta.com/tools/system-cards/multimodal-generative-ai-systems/.

[65] Microsoft. 2024. Copilot. Retrieved 12 April, 2024 from https://copilot.microsoft.com/.

[66] Omid Mohaddesi, Noah Chicoine, Min Gong, Ozlem Ergun, Jacqueline Griffin, David Kaeli, Stacy Marsella, and Casper Harteveld. 2023. Thought Bubbles: A Proxy into Players' Mental Model Development. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. ACM, Article 729, 21 pages. https://doi.org/10.1145/3544548.3581570

[67] Lourdes M. Morales, Sonia M. Arteaga, Peter Cottrell, and Sri Kurniawan. 2014. Design Guidelines of Tools for Facilitating Blind People to Independently Format Their Documents. In *Computers Helping People with Special Needs*, Klaus Miesenberger, Deborah Fels, Dominique Archambault, Petr Peňáz, and Wolfgang Zagler (Eds.). Springer International Publishing, Cham, 634–641.

[68] Cecily Morrison, Martin Grayson, Rita Faia Marques, Daniela Massiceti, Camilla Longden, Linda Wen, and Edward Cutrell. 2023. Understanding Personalized Accessibility through Teachable AI: Designing and Evaluating Find My Things for People who are Blind or Low Vision. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility* (New York, NY, USA) *(ASSETS '23)*. ACM, Article 31, 12 pages. https://doi.org/10.1145/3597638.3608395

[69] Fatemeh Mosaiyebzadeh, Seyedamin Pouriyeh, Reza Parizi, Nasrin Dehbozorgi, Mohsen Dorodchi, and Daniel Macêdo Batista. 2023. Exploring the Role of ChatGPT in Education: Applications and Challenges. In *Proceedings of the 24th Annual Conference on Information Technology Education* (Marietta, GA, USA) *(SIGITE '23)*. ACM, 84–89. https://doi.org/10.1145/3585059.3611445

[70] Dan Myers and Brian P Blake. 2017. An Evaluation of the Conformed Dimensions of Data Quality in Application to an Existing Information Quality-Privacy-Trust Research Framework. (2017). https://ualr.edu/informationquality/files/2023/02/P26-ICIQ2017-Conformed-Dimenstions-of-DQ.pdf

[71] Donald A. Norman. 2014. Some Observations on Mental Models. In *Mental models*. Psychology Press, 15–22.

[72] National Federation of the Blind. 2024. Blindness Statistics. Retrieved 23 April, 2024 from https://nfb.org/resources/blindness-statistics.

[73] OpenAI. 2024. ChatGPT. Retrieved 12 April, 2024 from https://openai.com/chatgpt.

[74] Eng Lieh Ouh, Benjamin Kok Siew Gan, Kyong Jin Shim, and Swavek Wlodkowski. 2023. ChatGPT, Can You Generate Solutions for my Coding Exercises? An Evaluation on its Effectiveness in an Undergraduate Java Programming Course. (2023), 54–60. https://doi.org/10.1145/3587102.3588794

[75] Konstantinos S. Papadopoulos, Evmorfia K. Arvaniti, Despina I. Dimitriadi, Vasiliki G. Gkoutsioudi, and Christina I. Zantali. 2009. Spelling performance of visually impaired adults. *British Journal of Visual Impairment* 27, 1 (2009), 49–64. https://doi.org/10.1177/0264619608097746

[76] Alisha Pradhan, Kanika Mehta, and Leah Findlater. 2018. "Accessibility Came by Accident": Use of Voice-Controlled Intelligent Personal Assistants by People with Disabilities. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. ACM, 1–13. https://doi.org/10.1145/3173574.3174033

[77] David Redmond. 2023. This Face Looks Redacted: The Be My Eyes Faces Controversy Explained. Retrieved 23 April, 2024 from https://vi.ie/this-face-looks-redacted-the-be-my-eyes-faces-controversy-explained/.

[78] Matthew Rueben, Jeffrey Klow, Madelyn Duer, Eric Zimmerman, Jennifer Piacentini, Madison Browning, Frank J. Bernieri, Cindy M. Grimm, and William D. Smart. 2021. Mental Models of a Mobile Shoe Rack: Exploratory Findings from a Long-term In-the-Wild Study. *ACM Transactions on Human-Robot Interaction (THRI)* 10, 2, Article 16 (feb 2021), 36 pages. https://doi.org/10.1145/3442620

[79] Heleen Rutjes, Martijn Willemsen, and Wijnand IJsselsteijn. 2019. Considerations on Explainable AI and Users' Mental Models. In *CHI 2019 Workshop: Where is the Human? Bridging the Gap Between AI and HCI*. ACM, Inc. https://www.martijnwillemsen.nl/recommenderlab/RutjesChHI2019ws.pdf

[80] Abir Saha and Anne Marie Piper. 2020. Understanding Audio Production Practices of People with Vision Impairments. In *Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility* (Virtual Event) *(ASSETS '20)*. ACM, Article 36, 13 pages. https://doi.org/10.1145/3373625.3416993

[81] Filipo Sharevski and Aziz N Zeidieh. 2023. "I Just Didn't Notice It:" Experiences with Misinformation Warnings on Social Media amongst Users Who Are Low Vision or Blind. In *Proceedings of the 2023 New Security Paradigms Workshop* (Segovia, Spain) *(NSPW '23)*. ACM, 17–33. https://doi.org/10.1145/3633500.3633502

[82] Renee Shelby, Shalaleh Rismani, Kathryn Henne, AJung Moon, Negar Rostamzadeh, Paul Nicholas, N'Mah Yilla-Akbari, Jess Gallegos, Andrew Smart, Emilio Garcia, and Gurleen Virk. 2023. Sociotechnical Harms of Algorithmic Systems: Scoping a Taxonomy for Harm Reduction. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society* (Montréal, QC, Canada) *(AIES '23)*. ACM, 723–741. https://doi.org/10.1145/3600211.3604673

[83] Marita Skjuve, Asbjørn Følstad, and Petter Bae Brandtzaeg. 2023. The User Experience of ChatGPT: Findings from a Questionnaire Study of Early Users. In *Proceedings of the 5th International Conference on Conversational User Interfaces* (Eindhoven, Netherlands) *(CUI '23)*. ACM, Article 2, 10 pages. https://doi.org/10.1145/3571884.3597144

[84] Yuan Sun, Eunchae Jang, Fenglong Ma, and Ting Wang. 2024. Generative AI in the Wild: Prospects, Challenges, and Strategies. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. ACM, Article 747, 16 pages. https://doi.org/10.1145/3613904.3642160

[85] David Touretzky, Christina Gardner-McCune, and Deborah Seehorn. 2023. Machine Learning and the Five Big Ideas in AI. *International Journal of Artificial Intelligence in Education* 33 (2023), 233–266. https://doi.org/10.1007/s40593-022-00314-1

[86] Stephanie Valencia, Richard Cave, Krystal Kallarackal, Katie Seaver, Michael Terry, and Shaun K. Kane. 2023. "The less I type, the better": How AI-Language Models can Enhance or Impede Communication for AAC Users. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. ACM, Article 830, 14 pages. https://doi.org/10.1145/3544548.3581560

[87] Sheena Vasani. 2023. Be My Eyes AI offers GPT-4-powered support for blind Microsoft customers. Retrieved 23 April, 2024 from https://www.theverge.com/2023/11/15/23962709/microsoft-blind-users-open-ai-chatgpt-4-be-my-eyes.

[88] Mathias Peter Verheijden and Mathias Funk. 2023. Collaborative Diffusion: Boosting Designerly Co-Creation with Generative AI. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI EA '23)*. ACM, Article 73, 8 pages. https://doi.org/10.1145/3544549.3585680

[89] Markel Vigo and Simon Harper. 2013. Coping Tactics Employed by Visually Disabled Users on the Web. *International Journal of Human-Computer Studies* 71, 11 (November 2013), 1013–1025. https://doi.org/10.1016/j.ijhcs.2013.08.002

[90] World Wide Web Consortium (W3C). 2024. WCAG 2 Overview. Retrieved 21 April, 2024 from https://www.w3.org/WAI/standards-guidelines/wcag/.

[91] Jingran Wang, Yi Liu, Peigong Li, Zhenxing Lin, Stavros Sindakis, and Sakshi Aggarwal. 2023. Overview of Data Quality: Examining the Dimensions, Antecedents, and Impacts of Data Quality. *Journal of the Knowledge Economy* (2023), 1–20. https://doi.org/10.1007/s13132-022-01096-6

[92] Ruotong Wang, Ruijia Cheng, Denae Ford, and Thomas Zimmermann. 2024. Investigating and Designing for Trust in AI-powered Code Generation Tools. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency* (Rio de Janeiro, Brazil) *(FAccT '24)*. ACM, 1475–1493. https://doi.org/10.1145/3630106.3658984

[93] Yunlong Wang, Shuyuan Shen, and Brian Y Lim. 2023. RePrompt: Automatic Prompt Editing to Refine AI-Generative Art Towards Precise Expressions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. ACM, Article 22, 29 pages. https://doi.org/10.1145/3544548.3581402

[94] Yuxia Wang, Minghan Wang, Muhammad Arslan Manzoor, Fei Liu, Georgi Georgiev, Rocktim Jyoti Das, and Preslav Nakov. 2024. Factuality of Large Language Models in the Year 2024. *arXiv* (2024). https://doi.org/10.48550/arXiv.2402.02420 arXiv:2402.02420

[95] Laura Weidinger, Jonathan Uesato, Maribeth Rauh, Conor Griffin, Po-Sen Huang, John Mellor, Amelia Glaese, Myra Cheng, Borja Balle, Atoosa Kasirzadeh, Courtney Biles, Sasha Brown, Zac Kenton, Will Hawkins, Tom Stepleton, Abeba Birhane, Lisa Anne Hendricks, Laura Rimell, William Isaac, Julia Haas, Sean Legassick, Geoffrey Irving, and Iason Gabriel. 2022. Taxonomy of Risks posed by Language Models. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (Seoul, Republic of Korea) *(FAccT '22)*. ACM, 214–229. https://doi.org/10.1145/3531146.3533088

[96] Meredith Whittaker, Meryl Alper, Cynthia L. Bennett, Sara Hendren, Elizabeth Kaziunas, Mara Mills, Meredith Ringel Morris, Joy Lisi Rankin, Emily Rogers, Marcel Salas, and Sarah Myers West. 2019. *Disability, Bias & AI Report*. Technical Report. AI Now Institute. https://ainowinstitute.org/publication/disabilitybiasai-2019

[97] J.D. Zamfirescu-Pereira, Richmond Y. Wong, Bjoern Hartmann, and Qian Yang. 2023. Why Johnny Can't Prompt: How Non-AI Experts Try (and Fail) to Design LLM Prompts. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. ACM, Article 437, 21 pages. https://doi.org/10.1145/3544548.3581388

[98] Zhiping Zhang, Michelle Jia, Hao-Ping (Hank) Lee, Bingsheng Yao, Sauvik Das, Ada Lerner, Dakuo Wang, and Tianshi Li. 2024. "It's a Fair Game", or Is It? Examining How Users Navigate Disclosure Risks and Benefits When Using LLM-Based Conversational Agents. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. ACM, Article 156, 26 pages. https://doi.org/10.1145/3613904.3642385

[99] Yong Zheng. 2023. ChatGPT for Teaching and Learning: An Experience from Data Science Education. In *Proceedings of the 24th Annual Conference on Information Technology Education* (Marietta, GA, USA) *(SIGITE '23)*. ACM, 66–72. https://doi.org/10.1145/3585059.3611431

[100] Jiawei Zhou, Yixuan Zhang, Qianni Luo, Andrea G Parker, and Munmun De Choudhury. 2023. Synthetic Lies: Understanding AI-Generated Misinformation and Evaluating Algorithmic and Human Solutions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. ACM, Article 436, 20 pages. https://doi.org/10.1145/3544548.3581318