RESEARCH-ARTICLE

Accessibility of Profile Pictures: Alt Text and Beyond to Express Identity Online

MARTEZ E MOTT, Microsoft Research, Redmond, WA, United States

JOHN TANG, Microsoft Research, Redmond, WA, United States

EDWARD CUTRELL, Microsoft Research, Redmond, WA, United States

Accessibility of Profile Pictures: Alt Text and Beyond to Express Identity Online

Martez E. Mott
Microsoft Research, Redmond,
Washington, USA
mamott@microsoft.com

John Tang
Microsoft Research, Mountain View,
California, USA
johntang@microsoft.com

Edward Cutrell
Microsoft Research, Redmond,
Washington, USA
cutrell@microsoft.com

ABSTRACT

Profile pictures can convey rich social signals that are often inaccessible to blind and low vision screen reader users. Although there have been efforts to understand screen reader users' preferences for alternative (alt) text descriptions when encountering images online, profile pictures evoke distinct information needs. We conducted semi-structured interviews with 16 screen reader users to understand their preferences for various styles of profile picture image descriptions in different social contexts. We also interviewed seven sighted individuals to explore their thoughts on authoring alt text for profile pictures. Our findings suggest that detailed image descriptions and user narrated alt text can provide screen reader users enjoyable and informative experiences when exploring profile pictures. We also identified mismatches between how sighted individuals would author alt text with what screen reader users prefer to know about profile pictures. We discuss the implications of our findings for social applications that support profile pictures.

CCS CONCEPTS

• **Human-centered computing**; • **Empirical studies in accessibility**;

KEYWORDS

Accessibility, image description, screen reader, blind, profile pictures, avatars

1 INTRODUCTION

Profile pictures are common on social networking platforms such as Facebook and LinkedIn, and on various websites and services used for professional, educational, and recreational purposes. The content of profile pictures depends on numerous factors, including users' personal preferences and the social norms of the platform where the images reside. These images often create first impressions for users viewing profiles of friends, colleagues, family members, and strangers. Ultimately, profile pictures serve as a means of self-expression that allows users to convey aspects of their personalities, interests, and identities in social settings.

Profile pictures, however, like other forms of visual content, can be inaccessible to blind and low vision people who use screen readers. Although social platforms mostly now provide interfaces for authoring alt text for pictures posted on their sites, only a few platforms allow users to provide alt text for profile pictures. Consequently, when screen readers encounter profile pictures, they typically can only announce "image", perhaps with users' names, which does not convey any of the identifying visual information contained in the profile picture. Profile pictures are often a crucial part of forming first impressions of people you meet online (especially, for example, in dating applications or searching for job candidates [1, 13]). Additionally, profile pictures can be very useful in disambiguating multiple user accounts with the same name, enabling people to recognize which account belongs to the specific user they were looking for. Without alt text for profile pictures, blind and low vision screen reader users may have difficulty accomplishing these tasks that are routinely performed when interacting with people online.

Social platforms must consider how they will support users when authoring alt text for profile pictures. Although there are guidelines [36, 37] that provide best practices for authoring alt text descriptions, researchers have found that these practices are often poorly suited for describing people who appear in images due to the complexities of representing gender, race, and disability [3]. Researchers have also noted that one-size-fits-all image descriptions are inadequate for addressing the information preferences of blind and low vision people, which necessitates the need for contextually relevant image descriptions [27, 28]. Context can be particularly important when determining how to describe images of people, as research by Bennet et al. [3] identified six contexts where additional appearance descriptions could be beneficial. Due to the vital role alt texts play in providing access to visual content, and the importance of profile pictures as expressions of users' identities online, it is important to further our understanding of how context impacts what screen reader users want to know about profile pictures, and how they prefer that information to be conveyed.

It is also important to better understand sighted individuals' opinions on alt text for profile pictures. Sighted people might author alt text, but many sighted people may not know what blind and low vision screen reader users want in alt text descriptions of profile pictures. Prior work has explored ways to guide users through authoring alt text in different contexts (e.g., memes [8], graphs

[20]), but the uniqueness of profile pictures as forms of identity expression might require new approaches to facilitate the authoring of high-quality alt text.

Our research builds upon and extends prior research efforts to understand the complex nuances of image descriptions of people. We present findings from two semi-structured interview studies. In our first study, 16 blind and low vision participants explored a website with several fictional user profiles that included a profile picture and an accompanying image description. To understand how context influenced our participants' information preferences, we presented them with three different scenarios. Participants were asked to imagine that they were looking to: 1) hire for a customer service position; 2) network with employees at different companies; and 3) socialize with members of a video game forum. Photographs of people were used in the profiles for the hiring and networking scenarios, and renderings of human and non-human avatars were used in the video game forum scenario. In each scenario, participants experienced four different styles of image descriptions (default, brief, descriptive, and narrated recordings of the description), which served as probes [31] to elicit thoughts on description preferences. In our second study, seven sighted participants explored a variant of the website used by our screen reader participants. Instead of profile pictures, sighted participants were provided buttons that played an audio recording of a screen reader reading the image descriptions associated with each profiles' picture. Sighted participants experienced all four image description styles for profile pictures with people and avatars.

Our paper makes the following contributions: First, we present findings from our interviews with screen reader users, which showed that they preferred more detailed image descriptions, as the additional information allowed them to make more informed decisions when completing tasks. Also, the inclusion of narrated image descriptions as an alternative for text-based descriptions provides opportunities to explore new forms of descriptions for profile pictures. Second, we describe the perspectives sighted individuals have on what should be included in alt text for profile pictures. These findings demonstrate the lack of knowledge sighted individuals have about what screen reader users would like to know about profile pictures. Finally, we provide preliminary discussion on the challenges and opportunities associated with describing avatars and their varying illustration styles.

2 RELATED WORK

Our research is motivated by and builds upon prior efforts to inform accessible image description practices. In this section, we provide background on the importance and use of profile pictures, non-visual image description practices, and the information needs of screen reader users when interacting with visual content.

2.1 Presentation of Self in Profile Pictures

It is common for users of social platforms, which include social networking sites (e.g., Facebook and Twitter), video game communities (e.g., Xbox Live and PlayStation Network), and business applications (e.g., Slack and Microsoft 365), to upload or select profile pictures when creating accounts. Fundamentally, a profile picture is a form of self-presentation, which is the process in which people guide

and control the impressions others form of them [10]. Decisions people make about which profile pictures to display depend heavily on how they wish to present themselves in various social settings and contexts, and the potential benefits these presentations might afford. For example, Hancock and Toma found that members of online dating sites were likely to select images that enhanced their physical attractiveness but would not be seen as deceptive during in-person encounters [13]. Ford et al. [7] found that profile pictures were crucial in building relationships and establishing trust in developer communities on GitHub.

Normative profile picture behaviors can also emerge on specific sites and applications. In a meta-analysis of college students' Facebook profile pictures, Hum et al. [15] found that most images were inactive (82%), posed (76%), and included only the user in the photograph (42%). Thomson and Greenwood [29] explored patterns in profile pictures across several social networking sites and found considerable differences in their participants' usage and perception of profile pictures on different platforms. Most participants (61.7%) used separate profile pictures for each platform, and the authors identified differences in expression, subject of depiction, and amount of body shown across different sites. For example, participants were more likely to pose by themselves on Instagram and with other people on Facebook, and to use photos from the waist-up on Twitter and head-and-shoulders only photos on Instagram. Culture can also result in varied profile picture presentations. In their examination on differences between Chinese and American social media users, Zhao et al. [34] found that Chinese users were more likely to customize their profile pictures while American users were more likely to have profile pictures that included multiple people.

Findings from these studies demonstrate the important role profile pictures play in mediating our presentation of self in social contexts. Profile pictures without image descriptions—or with poorly constructed image descriptions—prevent screen reader users from easily accessing these presentations, which may limit social interactions on these platforms. Screen readers users may also encounter difficulties understanding norms that emerge around profile picture use, which could be detrimental in contexts where judgements are made based on first impressions provided by profile pictures. For example, Baert [1] found that the quality of profile pictures on Facebook could significantly influence who job recruiters invited to participate in interviews, and Reiss et al. [24] found that people can reasonably predict a person's level of educational achievement based on their profile picture. Researchers have also used profile pictures to categorize users' personalities using the big 5 personality traits (e.g., openness, extraversion, etc.) [17].

2.2 Approaches for Providing Non-visual Image Descriptions

Screen reader users typically interact with images through screen reader software that reads the alternative (alt) text that accompanies images. Over the years there has been increased interest in understanding how to establish guidelines and methods that provide screen reader users access to the visual content present in images. The Web Content Accessibility Guidelines offered by the World Wide Web Consortium (W3C) instructs content creators to provide

text alternatives for non-text content, but the instructions are general and lack the specificity that can be helpful when creating an image description [36]. For example, the guidelines propose that alt text for "informative images", like photos and illustrations, "should be at least a short description conveying the essential information presented by the image" [36]. The National Center for Accessible Media (NCAM) provides general guidelines for alt text, such as consider your audience, be concise, and be objective, as well as specific description guidelines for different types of visual media, like maps, diagrams, and photographs [37]. Researchers have proposed strategies that build upon and improve the use of these guidelines in practice. Morash et al. [20] studied how two methods for creating alt text affected the quality of image descriptions composed by novice authors and found that a template-based approach was more effective than providing the guidelines alone. Mack et al. [18] built a prototype interface for constructing alt text in the context of Microsoft PowerPoint and found that interfaces that provided suggestions on what to include in an image description resulted in higher quality alt text. Morris et al. [21] proposed a taxonomy that could be used to represent visual content non-visually, which included alternative representations such as music and kinesthetic feedback.

In addition to human-based approaches, there have been efforts to develop automatic solutions for creating image descriptions. Wu et al. [32] designed and deployed an automatic alt text system for Facebook that used computer vision techniques to identify salient themes (e.g., people, objects, activities) in images to compose image descriptions. Gleason et al. [9] created Twitter A11y, a browser extension that uses optical character recognition and other automated methods, like Caption Crawler [12], to provide alt text for images on Twitter. Seeing AI [38] is a service from Microsoft that uses computer vision technologies to describe images, recognize people, and to perform other identification activities like reading currency.

2.3 Information Needs of Screen Reader Users

Including alt text is an important step towards providing screen reader users access to visual content, but it is also important to consider the information screen reader users would prefer to have when engaging with image descriptions. MacLeod et al. [19] investigated blind people's experiences with social media images and found that the phrasing of captions provided by automatic systems that introduced uncertainty (e.g., "I'm not really sure . . .") influenced their participants' trust of the caption, which suggest that image description framing is relevant information to provide. Petrie et al. [23] interviewed five blind participants about what information they wanted to know about images they encountered on the web and found that purpose, emotion, location, colors, activity, and objects (e.g., buildings, people, etc.) were the most important elements. In an investigation to close the gap between what is provided in image descriptions and what users want included, Stangl et al. [27] interviewed 28 blind and low vision people and found elements that were desired universally, which included information about people, objects, and scene descriptions.

Findings from the studies conducted by Petrie et al. [23] and Stangl et al. [27] also conclude that where an image appears can affect how a person interprets the image description. In their investigation of how context might impact image description needs, Stangl et al. [28] found that information needs can be similar but also considerably different for an image that appears across multiple contexts. For example, when an image containing multiple people in a park was placed within the context of visiting a news site, participants wanted to know the attributes of the park, the attributes of the people, and the activity of the people. When that same image was placed within the context of a social networking site, participants wanted to know the experience of the people and the relationship between them. Other forms of context can also be important for influencing information needs. Bennett et al. [3] found that image descriptions of people that included descriptions of race, gender, and disability could be important in six scenarios; *avatar creation; encountering unknown people; during discussions about identity and appearance; when seeking to read a room and find community; learning representation in media; and seeking specific perspectives.*

2.4 Summary

Our work builds upon prior research by examining screen reader users' information preferences when exploring profile pictures in three contexts. We also elicited thoughts from sighted individuals on what they think should be included in alt text of profile pictures. Although prior work has explored the role of context when creating image descriptions, there are numerous contexts and tasks that remain to be explored that might present unique challenges and opportunities. We specifically elicited thoughts on profile picture descriptions in professional and networking contexts. There are also opportunities to expand our understanding of how different presentations of profile picture descriptions could impact engagement on social platforms. Our introduction of narrated descriptions suggests that they could potentially facilitate better social interactions by providing first impressions to screen reader users. Finally, we provide preliminary discussion on considerations for describing avatars, as human and non-human avatars may be used as profile pictures more commonly in both recreational and professional contexts.

3 METHOD: SCREEN READER USER INTERVIEWS

We conducted semi-structured interviews with 16 screen reader users to understand their information needs and preferences when interacting with profile pictures in three scenarios: hiring an employee, networking with people at other companies, and meeting members of a video game forum. We chose these scenarios because they represent unique activities that provide insight into screen reader users' preferences and behaviors for professional and recreational tasks. The hiring and networking scenarios provided a lens for understanding behaviors and preferences when performing different tasks on the same platform, and the video game forum scenario explored preferences around descriptions of human and non-human avatars. In each scenario, participants experienced four styles of image descriptions: default, brief, descriptive, and narrative. The goal of our study was to identify if and how the information participants desired depended on the contexts of the separate tasks.

Table 1: Samples of reference profile pictures and their image descriptions in the four information

Reference Profile Picture	Default	Brief	Descriptive	Narrated
	Erica Brown.	A woman with brown eyes and black hair.	A Black woman with brown eyes and black hair wearing red glasses, hoop earrings, a black t-shirt and holding a white cane while standing in the lobby of an office building.	"I'm Erica Brown, a Black woman with brown eyes and black hair wearing red glasses, hoop earrings, a black t-shirt and holding a white cane while standing in the lobby of an office building."
	Nicholas Grey.	A man with black hair.	A white man with black hair with a stubble beard wearing glasses with red-tinted lenses, a gray hat, and a black jacket.	"I'm Nicholas Grey, a white man with black hair with a stubble beard wearing glasses with red-tinted lenses, a gray hat, and a black jacket."
	Bryan Walker.	A man with brown eyes.	A white male police officer with brown eyes and a grey mustache wearing a blue uniform.	"I'm Bryan Walker, my avatar is a white male police officer with brown eyes and a grey mustache wearing a blue uniform."
	Michelle Lewis.	A red fox with brown eyes and a brown nose.	A red fox with brown eyes, a brown nose, and brown highlights wearing green goggles with a slight smile and a yellow tail peaking from behind.	"I'm Michele Lewis, my avatar is red fox with brown eyes, a brown nose, and brown highlights wearing green goggles with a slight smile and a yellow tail peaking from behind."

3.1 Participants

We recruited and interviewed 16 screen reader users (nine men, seven women, average age of 34.9 years, *SD*=8.8 years) to participate in our study. Fourteen of our participants lost their vision at birth or during childhood, with two participants reporting losing their vision during their 20s. We recruited participants with the assistance of two agencies that specialize in providing services to blind and low vision individuals. Participants had to be at least 18 years of age, reside in the United States of America, and identify as a blind or low vision screen reader user. All interviews were conducted remotely using Zoom, which allowed participants to use their preferred screen readers (e.g., VoiceOver, JAWS, etc.) when interacting with a website we created and shared with participants during the study. One to four researchers were present for each interview. One member of the research team led the interview while other members were present for note taking and to occasionally ask follow-up questions. Each interview lasted approximately 90 minutes and we provided a $75 Amazon gift card as a gratuity to each participant.

3.2 Study Materials

We created a website to allow participants to explore user profiles and the image descriptions accompanying the profile pictures. Samples of profile pictures and their accompanying descriptions are shown in Table 1. The website consisted of 12 pages, with each page containing six fictional user profiles. Profiles for the hiring

and networking scenarios each had a profile picture, name, job title, and location (city and state). Profiles for the video game forum scenario also had six fictional user profiles, which included a profile picture, name, and role within the community (e.g., moderator, guest, member). Unlike the hiring and networking scenarios, the profile pictures were not photos of real people, but were instead renderings of human and non-human avatars. We deliberately chose a diverse set of profile pictures with regard to gender, race, and disability for the hiring and networking scenarios. For the video game forum scenario, we wanted a mix of human and nonhuman avatars, as well as cartoonish and realistic human avatar depictions. The research team collaboratively and iteratively generated image descriptions in each image description style for each profile picture. A button at the bottom of each page (except for the final page) allowed participants to progress through the website. While we used the reference image to guide the development of all the image descriptions, the study participants could only perceive the profiles through the image descriptions.

3.3 Interview Protocol

Our interview protocol was designed to elicit thoughts and opinions from our participants' prior experiences with profile picture descriptions, and to encourage their reflections on what information they would prefer in the three scenarios we presented. Our interviews were conducted in five parts. First, we asked participants to describe their experiences with image descriptions of profile pictures and other images that appear on various social media sites. We asked

participants about the frequency in which they encountered image descriptions and to describe the quality of descriptions they encountered. Next, we introduced either the hiring or networking scenario by reading a script that described the task the participants were to imagine themselves performing. For the hiring task, participants were asked to imagine that they were managers seeking to hire an employee to fill an open customer service position and that they had performed a search on LinkedIn to find potential candidates. For the networking task, participants were asked to imagine that they were employees at a large retail company and were looking to network with employees at different companies in anticipation of finding a new job. After describing the initial scenario (either hiring or networking), interviewers shared the study website URL with each participant through the Zoom chat.

During each scenario participants explored four pages on the website, with each page containing the same profiles but with a different image description for the profile pictures. We created four distinct image description styles. The first image description was named *default*, which was simply the name on the user's profile. This style was implemented to mirror the experience of screen reader users who encounter profile pictures that do not contain alt text, either because it is not supported by the platform or because an image description was not provided by the user. Sites like LinkedIn currently just read out the name for profile pictures. The second image description was *brief*, which contained gender, hair color, and eye color (similar to information found on a driver's license). Third was *descriptive*, which included gender, race, hair color, eye color, type and color of clothing, pose (standing or sitting), and background detail (when appropriate). Finally, the *narrative* style was a spoken version of the descriptive image description read aloud by a voice actor that identified as having the same gender and race of the owner of the fictional profile. It is important to note that the narrative style included a first-person introduction (e.g., "I'm Erica Brown") that was not present in the brief or descriptive information styles (see Table 1). These information styles were presented in the same order in each scenario: default → brief → descriptive → narrative. Participants progressed through the four information styles in the same order and without any explanation of the style. We chose to present the image description styles in this order so that as participants progressed through the scenario more information about the profile pictures would be revealed, creating opportunities for reflection on what their experiences would be like if only that style was available to them.

We asked participants to reflect on their experiences with the different information styles after they finished exploring all four styles, which included questions on what they liked and disliked about the image descriptions and what information, if any, they would like to see included or removed. We counterbalanced the presentation of the hiring and networking scenarios because we wanted fresh perspectives on how the image description styles influenced participants' information needs. When presented with either the hiring or networking scenario at the beginning, participants could provide feedback that was unencumbered by their experiences with the other scenario.

After experiencing the hiring and networking scenarios, participants were asked to imagine that they were new to a community of video game enthusiasts and were looking to meet new members.

During the video game forum scenario participants experienced the same image description styles presented in the hiring and networking scenario, and in the same order: default → brief → descriptive → narrative. After exploring all the image description styles, we asked participants to provide feedback on the quality of the descriptions and how the different styles would impact their behaviors in the scenario. We always presented the video game forum scenario last because the image descriptions used in this scenario varied greatly from those used in the hiring and networking scenarios and we did not want to accidentally mislead participants in what to expect in the other scenarios. To conclude the study, we conducted a participant debrief to elicit any remaining thoughts and opinions. We also asked participants to consider how widespread adoption of image descriptions would impact their experiences on various platforms, and if they could imagine any potential barriers that might prevent people from providing image descriptions for their profile pictures.

3.4 Analysis

We recorded and transcribed each interview. We analyzed our data using the qualitative method of reflexive thematic analysis [5]. One member of the research team analyzed notes taken during the interviews to find relevant concepts and performed inductive coding on the interview transcripts to develop an initial coding scheme. Next, the research team collectively iterated on the codes to create a final coding scheme. The primary author then performed a deductive coding step where the transcripts were reanalyzed to find instances of the agreed-upon codes. Finally, the research team identified and categorized relevant themes that describe participants' preferences and dislikes for different information styles, how the information styles would impact their experiences for each scenario, and the impact of authenticity when trusting the provenance of image descriptions.

4 FINDINGS: SCREEN READER USER INTERVIEWS

Our participants expressed nuanced opinions regarding their image description preferences when exploring user profiles. In this section, we describe observations that were elicited during opening conversations, and after participants experienced the *hiring*, *networking*, and *video game forum* scenarios. We describe participants' prior experience with alt text for avatars and profile pictures, information participants felt was relevant or unnecessary, how image descriptions styles would affect their behaviors while performing the tasks, how language provided in image descriptions could influence perceptions of authorship, and tensions between benefits of narrated content and considerations of time. We denote quotes from our 16 screen reader participants using SRU#.

4.1 Prior Experiences with Profile Picture Alt Text and Avatars

We started our interviews with discussions on participants' prior experiences with alt text for images on social platforms, including profile pictures. Our findings are consistent with existing research that found that screen reader users frequently encounter images with no alt text, or with alt text that appeared to be created through

automated methods [2, 19, 22, 30]. Our participants generally felt that it was better to have automated image descriptions than to have no description at all, but they also felt that the information provided in the descriptions was too generic to be particularly useful. For example, SRU1 mentioned that some descriptive information might be conveyed, but that there was no way to verify the validity of the description: *"Sometimes you may get a little quick description of the graphic with the person standing outside or something like that. It's like person smiling, glasses. . . usually you just don't know. And you really can't see. . .like there's no way to try to figure it out, but everything is just usually real vague."*

We also asked participants about their experiences selecting or building avatars. Seven participants had prior experience selecting, creating, or attempting to create avatars. Three participants mentioned positive experiences creating avatars on Facebook (SRU2) and Memoji (SRU4, SRU10). They noted that the creation process was made accessible by the inclusion of detailed labels on customization options. SRU2 described her experience with Facebook avatars: *"I feel like Facebook has made a lot of steps forward, especially with their new avatar feature. They're super descriptive when you're picking the components of like skin color and the hair color and facial structure of your avatar. . . so that's super helpful."* For SRU11, however, there were too many details to customize, and she was not interested in exploring the available options: *"I was going to create an avatar on Facebook, but it was just so many details like I got tired of it. I mean, it makes sense if you want all that detail. What color eyebrow? What color hair will you try?"*

4.2 Information Preferences

We designed our three scenarios to elicit thoughts and opinions from participants when performing real-world tasks that have not been explored in previous investigations that sought to understand how context impacts information needs. Overall, we found that most participants (n=14) preferred the *descriptive* and *narrative* information styles, as they appreciated the additional information they provided compared to the *default* and *brief* styles. We also found that how participants would potentially use the information presented in the descriptions would depend on the task they were performing.

In the *hiring* scenario, participants commented that having more access to visual content included in profile pictures would allow them to make more informed hiring decisions. For example, SRU3 noted that the appropriateness of how people dressed would be a factor in her decision-making process: *"I think if you're looking for a job, you're putting your face out there, your profile out there, you want that picture to match what you're looking for. You're not going to be showing up looking for someone to work customer service wearing a tank top and shorts. Probably not going to be somebody I want to hire."* SRU10 felt that as a hiring manager it would be important for her to access the same visual information available to sighted people. SRU8 made an interesting observation about how more descriptive information could help ensure proper accommodations for people with disabilities: *"For the gentleman in a wheelchair, that was good that was in the description, because when hiring, thinking of accommodations without him even asking for it."*

During the *networking* scenario, participants discussed the benefits of additional profile picture information when potentially

connecting with strangers for the first time. One benefit was the social signals about workplace culture that can be conveyed through profile pictures. For SRU4, more detailed image descriptions would allow her to better understand if the company had a diverse workforce: *"I think when it comes from looking at another company, and sort of the individuals that work there, this is when physical characteristics may come a little bit more relevant. Just simply because for myself, I would like to see if there's representation from diverse communities, especially the ones that I'm a part of."* A second benefit was the opportunity to identify mutual interests, which could be useful for starting conversations. SRU7 explained: *"If they have photos that show them doing certain things, or having a certain hobby, obviously that could lead to conversations, so that could be useful to have. Like if you notice something in the photo that indicates a shared interest, like a sports team, or if they're wearing a piece of clothing, or if they're standing in front of a certain monument."*

The *video game forum* scenario was an opportunity to explore how our participants responded to human and non-human avatar renderings. Unlike profile pictures of people, avatars can be completely fantastical creations. As a result, more details can help screen reader users understand what is represented in the images. Because avatars provide additional opportunities for freedom and expressivity compared to typical profile pictures, participants were eager to access visual information so they could gain a deeper appreciation of users' personalities. SRU2 described how the avatar descriptions would affect whom she would interact with on the forum: *"I think you get to see more of the personality of a person through their avatar. In this situation, I think it gives you information about who you would wanna talk to. Like I love animals, so I would totally want to talk to the girl who's a fox and sounds really cute. Or the guy who picked candy corn sounds silly and more fun."* Two participants (SRU10, SRU13) noted that some blind people might not be aware of the various visual styles avatars can take, which would necessitate the need for even more detailed image descriptions. SRU10 commented on how a description like "cartoon" does not evoke specific imagery: *"Because I don't experience the world visually, I don't always know exactly what it is. Like maybe that should be a given that cartoons kind of tend to have exaggerated features, now that I think about it. But that was not coming to mind when I read this description."* SRU13 explained that including very specific details could be particularly helpful, such as describing the shape of the candy corn avatar: *"I think it takes for granted that people know the shape of a candy corn. . . the thing that is really distinct about candy corn is its triangular shape or cone shape. So I think that would be helpful. . . anytime you can include some of those kinds of very distinct details I think it's good."*

Overall, our participants noted that more detailed image descriptions could provide access to social signals that would impact their behaviors when performing different tasks. Clues about a workplace's culture, the appearance of a job candidate, and the personality of forum members, are all examples of how detailed image descriptions of profile pictures would allow screen reader users to engage with other users in meaningful ways on social platforms.

4.3 Narrated Image Descriptions

Participants generally enjoyed the rich experience offered by the narrated image descriptions. In particular, participants commented

that hearing users' voices made the profiles feel more personal, and that it allowed them to create mental images of users. SRU12 explained how hearing a voice can be similar to seeing a person for the first time: *"I love that because oftentimes when you first see someone, it's an impression you know... For a visually impaired person, hearing someone speak, you get that same impression and I thought that was such a great touch. It's like, oh it's a real person. You know, it's not just a name on a screen."* Another benefit of the narrated image descriptions was that hearing users' voices could provide additional context for screen reader users when performing different tasks. For example, SRU5 mentioned how hearing different voices would provide insight into who you might connect with on social platforms: *"The impact from just listening to them and being like, oh he sounds, you know, chill and friendly, or she sounds a little bit stern or things like that... I don't know what it is, but it says something about how much I rely on that as a blind person and how rich a source of information I feel it can be."* During the video game forum scenario, SRU15 described how narrated descriptions would improve his interactions on the forum if it was text based, since he would not have opportunities to voice chat with other members: *"If I'm kind of doing this for fun and it's a text forum, I'm not able to talk to any of these people. It'd be kind of cool to imagine what they sound like when I'm reading their posts in text, interacting with them in text."*

The narrated image descriptions also surfaced questions around image description authorship. By lending their voice to narrate their image descriptions, participants felt more confident that the image descriptions were actually authored by the narrators rather than AI or some independent third-party. For SRU2, the narrated description conveyed a sense of ownership that was not present in the text descriptions read by the screen reader: *"The two major differences between the last two pages is the first one was written out, and I don't necessarily know who wrote it. It could have been them, but it was kind of really nice to hear them take ownership of it."* SRU5 mentioned how first-person language—which is only present in the narrative image description—would influence who she believed authored the image descriptions. SRU5 continued to explain that is important for descriptions to be authored by people in the photographs when describing race and disability: *"Maybe someone who's Indian would be preferred to be referred to as Indian rather than Asian, or you know, doesn't want to be referred to that way at all. I mean, I think if a written description said I'm a black woman holding a white cane in the lobby of a building, like that, that would feel better if it was like authentic that they'd done it."*

One drawback participants identified about the narrated descriptions was the time required to listen to the descriptions compared to the text-based image descriptions read by their screen readers. Considerations for time can be important depending on the task being performed. For example, during the hiring scenario, SRU7 mentioned how it could become burdensome to listen to too many narrated descriptions when evaluating potential job candidates: *"Just to hear someone voice, I guess is a nice thing to be able to do... But I don't know how many people I'm looking through. If like, let's say I only had 10 people, or I don't know less than 100, this audio would take forever and I wouldn't want to. I would just skip it altogether... it just depends on how many people I'm looking at."* The narrated image

descriptions also did not provide screen reader users access to typical screen reader interactions to navigate the image descriptions. SRU4 described why the controls afforded by her screen reader lead her to prefer text-based over narrated image descriptions: *"So if you wanted to listen to, you know, one word at a time, or go to specific word in the center and identify sort of a specific item that was mentioned, then you are able to navigate to that part of description without having to listen to the recording from the very beginning."*

4.4 Discrimination and Bias

Two participants (SRU1, SRU15) expressed concerns that detailed image descriptions could potentially lead to race- or gender-based discrimination. SRU15 enjoyed the descriptive information style but felt that race should have been excluded from the image description: *"I almost feel like race should be cut out because that creates potential room for discrimination. Whereas they don't feel like it's super relevant. But things like outfit, background, appearance, all that stuff is excellent."* SRU1 questioned if job candidates would intentionally leave out gender and race information in their image descriptions to avoid potential discrimination: *"Are people going to purposely leave things out of their picture? Would a woman with a gender-neutral name want to leave it out of the description because they think they're gonna get discriminated against? Same with someone, you know, of Asian descent. Or you know, African American, whatever."* SRU1 and SRU15 make great observations about the potential harms associated with users explicitly stating their race or gender, especially within the context of hiring.

5 METHOD: SIGHTED USER INTERVIEWS

After concluding the interviews with our screen reader participants, we were curious to know what opinions sighted individuals had regarding authoring image descriptions for profile pictures. As we mentioned previously, because sighted people are producers of image descriptions but not consumers of them—unlike screen reader users who both produce *and* consume image descriptions—it is important to understand their opinions on what to include in image descriptions.

5.1 Participants

We recruited and interviewed seven sighted individuals (six women, one man, average age of 37.1 years, *SD*=5.8 years) from a large technology company headquartered in the United States to participate in our study. Participants had to be at least 18 years of age, reside in the United States of America, and have previous experience with profile pictures on social platforms. Prior experience with alt text was not a requirement for participation, but three participants stated that they had previously written alt text for images. Interviews lasted approximately 60 minutes and were conducted remotely over Microsoft Teams with one lead interviewer and up to two research team members present for each interview. We provided each participant a $50 Amazon gift card as gratuity.

5.2 Study Materials

We created a second website based on the one used with our screen reader participants (see 3.2) to allow sighted participants to explore image descriptions of profile pictures of fictional user profiles. The

website comprised ten pages with six fictional profiles per page. The first five pages contained profiles from the *networking* and *hiring* scenarios (see 3.2) and the remaining pages used profiles from the *video game forum* scenario (see 3.2). We replaced the profile pictures with buttons that would play recordings of a screen reader reading the image description associated with each picture. Each page would play the image description associated with one of four image description styles (*default*, *brief*, *descriptive*, and *narrative*). The final page in each scenario (pages five and ten) presented the pictures for each profile as well as four buttons that would play the audio associated with each image description style.

5.3 Interview Protocol

We designed our interview protocol to prompt reflections from our sighted participants on how they think about authoring image descriptions for profile pictures. First, we asked participants to discuss their thoughts on and experiences with profile pictures on social platforms. Next, we asked participants to author image descriptions for profile pictures they currently or previously used. We then shared the study website URL with participants and read a script explaining the details found in the profiles, and that they were to listen to each profile's image description before proceeding to the next page.

Participants experienced image description styles in the same order as our screen reader participants: default → brief → descriptive → narrative. On the pages after the narrative image descriptions (pages five and ten), participants were shown the images for each profile, and they were provided buttons to play audio for each image description style. We asked participants to discuss their opinions on the different image description styles and to reflect on what information they felt was useful, missing, or unnecessary. Finally, after participants explored all the website pages, we asked them to think about the profile picture image description they authored earlier in the study and to consider how, if at all, they would change their description.

5.4 Analysis

We recorded and transcribed each interview. We used the qualitative method of reflexive thematic analysis [5] to analyze data from sighted participants. One member of the research team performed a deductive coding step where the transcripts were analyzed using the codes from our initial study with screen reader users, followed by an inductive coding step to identify additional codes. The research team then iterated and finalized the codes, and finally, we identified and grouped thematically relevant quotes. We identified themes concerning participants' misconceptions of what information screen reader users would prefer to have in image descriptions, the importance of credibility in narrated image descriptions, and the potential for bias when disclosing identify information.

6 FINDINGS: SIGHTED USER INTERVIEWS

We present observations from our interviews with sighted individuals, with a focus on what information our participants felt should be included in image descriptions of profile pictures. We also highlight participants' thoughts on narrated image descriptions, and

concerns about explicitly providing parts of their identity (e.g., gender, race) in descriptions. We denote quotes from our seven sighted participants using SP#.

6.1 Misconceptions about Profile Picture Image Description Preferences

All seven of our participants held misconceptions about what information screen reader users want to know in image descriptions of profile pictures. We found that our participants would often question if the information presented in an image description was necessary. Two areas where participants had the most difficulty understanding their importance were descriptions of backgrounds and of users' clothes and jewelry. SP7 explained that because many profile pictures are headshots, the background becomes less important: *"Because profile pictures are usually up close, the background is actually not important. So, I would not care as much about the background."* For SP6, descriptions of jewelry seemed unnecessary: *"So one thing that I felt could be a little redundant is like, do we really need to call out, you know, choker necklace or hoop earrings? Because we are giving too much details about hoop earrings and a choker. What does that add?"* These opinions are contrary to what we learned from our screen reader participants in Study 1, who described the benefits of these details when performing tasks. For example, in section 4.2 we described how SRU7 would use descriptions of clothing or backgrounds to potentially identify common interests with other users.

What participants thought was necessary information was often based on their own opinions and expectations, rather than what they thought might be useful to people who use screen readers. SP3 described how his personal preferences would influence the amount of detail he would provide in image descriptions: *"One of the things I felt like when I was hearing some of the descriptions was like, OK, it was more than what I needed. Maybe that's just the way that I think about things, but I think it was like, yeah, maybe there was a little too much detail, where even just like the basic thing would be good enough for me. Like there's a profile picture with someone in there and like that's all I needed to know."* Participants also expressed confusion about what items to include in image descriptions, and how much description they should provide on those items. For example, SP1 questioned how much detail is necessary when describing the glasses worn by one of the fictional users: *"Like I said with Katherine, those are big glasses. I would say that level of granularity doesn't matter to people. And since I don't have a visual disability, I don't use screen readers, I don't know how much they want those nuances like for her. That's almost like cat eyeglasses from the 60s."*

The avatar profile pictures also elicited questions about what information to include in image descriptions. Like their thoughts on profile pictures of humans, participants questioned if details about attire (or similar physical details) were necessary. When describing her feelings about the fox avatar, SP4 thought describing the fox's tail was too much: *". . .especially with like the yellow tail peeking from behind, you know, it's a lot of detail there."* SP6 expressed a similar sentiment when discussing an avatar's clothing: *"For example, the elf on the t-shirt didn't add much value to me."* Participants did note that because avatars are constructions rather than photographs of people, there are opportunities to provide more descriptions on

why users created or selected certain avatars. SP1 compared the process of creating or selecting an avatar to telling a story: *"It feels like they're, you know, telling a story. Sometimes creation is telling a story. It's like I picked an avatar that looks like me. I picked an avatar who doesn't look like me. I picked an avatar that's my favorite animal."*

Misconceptions held by our sighted participants were also found in the alt text they authored for their profile pictures. Six profile pictures chosen by our participants were headshot style photographs and one profile picture was a photo of a bovine, which the participant stated was used for preserving privacy but also expressing his quirky personality. When asked to initially write alt text for their profile pictures, four participants included their gender, three included their race or ethnicity, and one included a description of their background. However, after experiencing the different information styles, all participants with human profile pictures included both gender and race/ethnicity, and three included a description of their background. The biggest difference in alt text content was from SP6, who initially wrote her description as "*young, and lively, looking to go above and beyond*" and changed it to, "*Asian Indian women in her mid 20's, with a smile on her face, black hair, black eyes, leaning near a white wall and wearing a top with stripes.*" Overall, our sighted participants were either unsure what information to include in profile picture image descriptions, or disregarded information that could be beneficial to screen reader users. This mismatch between what sighted people think should be included in image descriptions and what screen reader users actually want is problematic, as sighted individuals will be responsible for authoring their own image descriptions.

6.2 Benefits of Narrated Image Descriptions

Our seven participants found the narrated image descriptions to be enjoyable and engaging, echoing many of the sentiments shared by our screen reader participants. Participants commented that hearing image description in users' own voices seemed more personal. SP4 felt that you could pick up on users' energy by hearing their voices: *"I love that descriptions are in the persons' own voices. I really like that. You know, just being able to have that. Like you pick up on a vibrant energy right when the person reads in their own voice, right?"* For SP5, hearing a person's voice might provide an opportunity to form a different opinion of them: *"I think allowing the voice could kind of help, cause if you just hear dreads, Black, or whatnot, it could go to a negative place with some people. So I think allowing the voice and the tone can kind of give a different type of visual to the person if they're not able to see him."*

Participants also thought that the narrated description contributed a sense of ownership that was not conveyed by the other image description styles. Although the narrated description was the only one to include a first-person introduction, our sighted participants focused less on the language and more on how contributing voices to profiles would increase the credibility of users' image descriptions. SP1 described how the narrated description reinforced her opinion that the image descriptions were what the users wanted to convey about their profile picture: *"To me, it was again that sense of ownership. Like when I look at all these different pictures, and I hear how they decided to describe themselves. It's like,*

this is how I want people to see me, and that's often what I think the profile picture is."

For avatars, participants were less enthusiastic about narrated descriptions. Participants had concerns that narrating an image description could potentially break the anonymity offered by avatars. SP2 explained: *"Because these are avatars, for some reason I didn't want the person who actually created them to narrate. I was more comfortable with the computer generated one. . .I think it's because of what an avatar represents. There's like some amount of abstraction in the virtual world. It's almost like you, but you're also kind of hiding yourself a bit."* Participants also felt that because avatars were representations of users, rather than the users themselves, it was unnecessary to hear their voices. SP7 mentioned that she was not expecting to hear narrated descriptions for the avatars: *"For some reason, I was thinking that if it's an avatar, I don't expect it to be personally narrated. . .Maybe because it's an avatar right, so you don't personally identify with it."*

6.3 Concerns about Identity Disclosure

In section 4.4 we described how some screen reader participants expressed concerns that people might be hesitant to disclose identity information about gender or race in their image descriptions. We found that three sighted participants (SP2, SP3, SP4) had explicit concerns about making identity information known in image descriptions. SP2 and SP4 were concerned that including information about gender, race, or other personal details, might lead to discrimination. SP2 described how she would be hesitant to mention that she was pregnant in an image description for an old profile picture: *"So you couldn't really tell, but if I had submitted that picture when I was like 8 months pregnant you would be able to sort of maybe tell that I was pregnant, depending on how they took the picture. Let's just say I uploaded it and then a screen reader ended up talking about it like, this woman is an Asian Indian woman wearing a gray sweater, unmasked, brown eyes and pregnant or something. Like that's not a thing you want to repeat. Especially on a place like LinkedIn, because then companies may not want to hire you, or they might realize, like, oh, she probably already had the baby and already went on parental leave. But like I don't know, there's all sorts of weird things about disclosing certain things, so that's where I was feeling weird about this."* SP3 expressed concern that people might be discriminated against based on their narrated image description: *"I feel especially mindful about bias on here. Like if someone had an accent or something, I feel like I would kind of have in the back of my mind, like am I favoring or not favoring someone because of the way that they speak?"*

These perspectives highlight the complicated nature of describing people in profile pictures. Although sighted users might make assumptions about people's gender, race, etc., authoring image descriptions requires explicit identity disclosure, which can be an uncomfortable or difficult task for some people.

7 DISCUSSION

Researchers and practitioners in the accessibility community have spent considerable effort discovering how to provide screen reader users access to visual information present in images [4, 6, 12, 14, 16, 18, 21, 23, 25, 27, 28, 35]. Our work builds upon and extends these prior efforts by focusing on the information screen reader users

prefer to know when encountering profile pictures, and on the perspectives sighted individuals have around what should be included in profile picture descriptions. The focus of our work was to expand our understanding of how the appearance of profile pictures within different contexts influences the information needs of screen reader users, ultimately so that applications and platforms can provide relevant image descriptions when and how users want them. Also, by understanding the misconceptions sighted individuals have regarding profile picture alt text, we can think about how to prompt users to author descriptions that meet the needs of screen reader users. In this section, we discuss the role of profile pictures as a form of identity representation; how context influenced information needs; the disjointed image description experiences between screen reader and sighted users; the potential of narrated image descriptions; and the unique image description challenges posed by avatars.

7.1 Identity Representation in Profile Pictures

Profile pictures perform the unique role of representing users' identities on social platforms, perhaps even more than other posted photos. Contributing to the importance of profile pictures is their prevalence throughout different parts of social applications. A profile picture can appear on a user's homepage, next to a user's comment, or in list after a search result. Because profile pictures are so ubiquitous, it makes sense that users would actively select profile pictures that represent aspects of their personalities and identities [10]. Selecting a profile picture is a *front-stage* behavior [10], where users intentionally select a photo to represent them in a specific context.

The importance of profile pictures as expressions of identity places an expectation that profile picture alt text can also successfully convey aspects of people's identities. Unlike photos, however, where aspects of identity must be interpreted by the viewer, alt text is unique in that users can explicitly express aspects of their identity in writing. As we found, both our screen reader and sighted participants expressed concern that people would be apprehensive about disclosing identity information in alt text, especially in contexts associated with hiring or job seeking. These findings surface new questions and concerns on whether people with marginalized identities would feel comfortable disclosing identity information in alt text for profile pictures in professional contexts. Bennet et al. [3] identified and discussed tensions surrounding disclosures of race, gender, and disability, and noted that people exhibited varied personal preferences for when to disclose identity information. The authors also surfaced concerns about the potential burden placed on marginalized people to disclose identity information, which is a point raised by SRU1 when he wondered if women and racial minorities would feel comfortable including aspects of their identities in image descriptions. SP2 also shared this view when expressing concern that she might be discriminated against if she were to disclose that she was pregnant. Although Bennet et al. suggested that additional appearance descriptions could be useful in the context of *encountering unknown people*, further research is needed to understand the thoughts and preferences of people with marginalized identities when disclosing identity information in contexts such as hiring.

We also need additional research to understand how platforms should encourage users to disclose identity information in image descriptions for profile pictures. Participants noted that identity information could be useful when trying to understand if a company has a diverse employee base, or when attempting to be proactive in accommodating accessibility needs. However, if users are uncomfortable or unwilling to provide identity information, screen reader users will miss opportunities to perform these activities.

7.2 The Impact of Context

Our findings show that screen reader users would like access to detailed information present in profile pictures, and our participants provided numerous examples to illustrate how they would use that information when completing different tasks. These findings are in line with prior work that has consistently shown that screen reader users typically prefer to have certain details conveyed in image descriptions [23, 27, 28], although the relevance of more specific information can be heavily influenced by the context of where the image appears [27, 28]. Within the context of hiring, participants mentioned wanting descriptions of clothing or jewelry so they could determine if job candidates were presenting themselves professionally. In particular, the style and type of clothing could be helpful, such as the distinction between a t-shirt and a collared shirt. Style of dress has been proposed as a detail of interest for images that appear on e-commerce sites [26] but our findings suggest that similar details might be beneficial on platforms like LinkedIn that are tailored toward hiring and job seeking.

For networking, participants described how details about attire and backgrounds could help them identify if they share common interests with other users. It could be valuable for alt text authors to describe details that might be traditionally overlooked, like providing the name of a book displayed in the background, or describing a souvenir purchased while traveling abroad. Of course, authors will have to manage the amount of information they provide—describing every book on a bookcase might be extreme—but supplying some description of objects that can be identified visually might provide screen reader users opportunities to find shared interests or hobbies with other users online.

7.3 Mismatches Between Screen Reader and Sighted Users

We found that our sighted participants were often unsure about what should be included in image descriptions, and frequently questioned the importance of various details. For example, the comment by SP7 that backgrounds were not important is in opposition to the opinions expressed by our screen reader participants, who found value in the social signals provided by descriptions of backgrounds. Overall, the tendency was for our sighted participants to suggest that less information be provided. An issue at play is that our sighted participants' views on what should be included in image descriptions were typically based on their personal opinions, not on what they thought might be useful to screen reader users. Mismatches between screen reader and sighted users is an example of Problem 1 of "Why CSCW applications fail" presented by Grudin [11]: *the disparity between who does the work and who gets the benefit.* This problem is often exacerbated in accessibility contexts since

the people who do the work often do not understand the needs or perspectives of the people who will benefit. Prior research has explored the effectiveness of various methods for authoring alt text for different types of imagery, including memes [8], charts [20], and images in PowerPoint [18]. Future research should explore the creation and evaluation of different approaches (templates, guidelines, etc.) for eliciting high-quality image descriptions for profile pictures.

7.4 Potential for Narrated Image Descriptions

Our narrated information style introduced a unique way for participants to engage with image descriptions that are typically read by a screen reader. In particular, the narrated style exhibited two fascinating traits: (1) the richness provided by audio; and (2) ownership of the profile established through first person language.

7.4.1 Richness of Audio. Both our screen reader and sighted participants thought the narrated descriptions were an engaging way to interact with profile pictures, which suggests that narrated image descriptions, or other narrated content, could have broad appeal [33]. Our screen reader participants noted how hearing users' voices would afford opportunities to learn more about personalities, or to have a more personal experience on social platforms. For example, SRU15 mentioned how hearing a user's voice from their image description would make reading their post more enjoyable, as he could imagine the words being spoken by the user. Our screen reader participants shared similar views on the benefits of narrated descriptions in both the *networking* and *video game forum* scenarios, which is unsurprising since both scenarios centered around connecting with new people. In these scenarios, the narrated descriptions conveyed personality traits that influenced participants' desires to make connections. As noted by SRU9, voices can serve as a kind of first impression for blind people, which presents an opportunity for narrated descriptions to fill this role on social platforms. Narrated descriptions can also potentially help screen reader users disambiguate between user accounts with similar profile details.

Our screen reader participants did comment that it took significantly longer to listen to the narrated descriptions compared to the text-based descriptions, since many screen reader users set the speaking rates of their screen reader software to speeds much faster than conversational speech. The length of the narrated descriptions could impact participants' behaviors in different contexts. For example, SRU7 noted how it would be best to skip the narrated descriptions if evaluating dozens or hundreds of job candidates. There are opportunities to improve the narrated description experience by allowing users to interact with narrated descriptions similar to how they interact with alt text. SR4 mentioned how his screen reader allows him to listen to one word at a time, or jump to a specific part of the text. Social platforms could provide similar functionalities for narrated image descriptions, which would afford screen reader users more freedom and flexibility in how they interact with image descriptions. Future work should explore how screen reader users engage with narrated descriptions in more natural settings to better understand how to design effective methods for interacting with narrated descriptions.

7.4.2 Profile Ownership. An important distinction between the *narrated* and *descriptive* information styles is that the *narrated* style included a first-person introduction that preceded the image description (e.g., "I'm Erica Brown, a Black woman..."). Including this introduction strongly implied ownership of the profile to the image description narrator. As noted by SRU2, it can be difficult or impossible to determine who authored a textual image description, which could result in a lack of trust. SRU5 echoed a similar sentiment by describing how it can feel invasive to have someone describe aspects of an identity, like race or disability, for another person. These comments share similarities with preferences shared by participants in Bennett et al.'s [3] study, who said that appearances should be described using the preferred language of the person in the photograph. The first-person introduction included in the audio description does provide a potential way for image description authors to establish trust by including first-person identifiers in the description. It is possible that by including phrases like "I am" could increase authenticity of image descriptions. A design implication is that templates and other approaches that advise people on how to compose image descriptions could include prompts that instruct authors to use first-person language when describing their own profile pictures. The inclusion of first-person language in both text and audio formats could result in more credibility with screen reader users. Further research is needed to better understand how and when first-person language or audio narration might impact screen reader users' trust of profile picture descriptions.

7.5 Considerations for Avatar Image Descriptions

The inclusion of human and non-human avatars in the *video game forum* scenario allowed us to make observations about the complexities of providing image descriptions of avatars.

7.5.1 Range of Styles and Representations. Our screen reader participants especially enjoyed the more detailed information descriptions of avatars, and it was particularly interesting to note how participants responded to differences in information presented between the *brief* and *descriptive* styles. Participants generally felt that the brief descriptions left out crucial information that would impact their socialization behaviors. For example, SRU0 was surprised to hear that Brian's avatar was wearing a police uniform, describing that detail as a *"a pretty big deal"*. Identifying and describing specific details seemed particularly important in this scenario, primarily because there was more variety in the type of image descriptions available. Poorly described avatar descriptions could have considerable negative effects on the social experiences of screen reader users, as information that participants found useful for socializing, like facial expressions and style of clothing, could be excluded. It is important to note that avatars are commonly used as profile pictures on entertainment platforms like Xbox and Netflix, and avatars are becoming more commonplace in professional settings with the emergence of virtual reality collaboration spaces like Meta's Horizon Workrooms.

7.5.2 Additional Information for Improved Perception. Our screen reader participants expressed concern that sighted people may be unaware of what blind people know about the visual characteristics

of certain objects. For example, SRU13 noted that the description of the candy corn avatar did not include information about its cone-like shape, which is a detail that might not be known by all screen reader users. In addition, language descriptions of illustration styles, such as "cartoon", might not be fully understood. SRU10 mentioned that since she has not experienced the world visually, she was unaware that cartoon images might have exaggerated features like big eyes. This is an interesting observation, because avatars might portray one of many different illustration styles that each possess their own distinct visual characteristics. Authors should provide additional details that can more accurately depict unique avatar characteristics, especially for illustration styles that convey specific visual information. Future work should explore screen reader user preferences when encountering profile pictures of avatars with varying illustration styles.

7.6 Limitations

We did not specifically recruit screen reader users that identified with marginalized gender or racial identities, nor did we ask participants to disclose their racial identities. It is possible that screen reader users with marginalized racial and gender identities, as well as intersecting disabilities, might have provided opinions or concerns that we were unable to obtain in our study. Also, since all our screen reader user and sighted participants were recruited from the United States, we could not gain perspective for how culture might influence screen reader users' thoughts and preferences on profile picture descriptions. We designed the four information styles and the corresponding image descriptions to help facilitate discussion during our study, and it is possible that different descriptions might have elicited responses not captured in our data. We explored three common real-world scenarios, but there are numerous and varied contexts where profile pictures might appear, such as online dating platforms, that could surface information needs not explored in our study. Also, our *hiring* scenario was not representative of all hiring activities that occur on professional networking sites that could necessitate different information needs, such as hiring a model or actor.

8 CONCLUSION

We presented findings from two semi-structured interview studies. Our first study with 16 screen reader users advanced our understanding of what information screen reader users would prefer to have when exploring profile pictures in different social contexts. We found that our screen reader participants generally enjoyed more detailed information about profile pictures, but that tensions surfaced when considering the disclosure of identity information. The inclusion of narrated image descriptions introduced new opportunities for screen reader users to form first impressions of people they encounter online. We also found that our screen reader participants appreciated having more details in the image descriptions of avatars, and that there are additional opportunities for conveying different avatar illustration styles. Our second study with seven sighted individuals highlighted the mismatch between what information sighted people think belongs in alt text, and what information screen reader users prefer to know. Finally, we presented design implications for establishing credibility by using

first-person language in image descriptions, and for developing more interactive experiences for narrated image descriptions.

REFERENCES

[1] Stijn Baert. 2018. Facebook profile picture appearance affects recruiters' first hiring decisions. *New Media & Society* 20, 3: 1220–1239. https://doi.org/10.1177/1461444816687294

[2] Cynthia L. Bennett, Jane E, Martez E. Mott, Edward Cutrell, and Meredith Ringel Morris. 2018. How teens with visual impairments take, edit, and share photos on social media. In *Proceedings of the ACM Conference on Human Factors in Computing Systems* (CHI '18), 76:1-76:12. https://doi.org/10.1145/3173574.3173650

[3] Cynthia L. Bennett, Cole Gleason, Morgan Klaus Scheuerman, Jeffrey P. Bigham, Anhong Guo, and Alexandra To. 2021. "It's Complicated": Negotiating Accessibility and (Mis)Representation in Image Descriptions of Race, Gender, and Disability. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 1–19. Retrieved November 3, 2021 from https://doi.org/10.1145/3411764.3445498

[4] Erin Brady, Meredith Ringel Morris, Yu Zhong, Samuel White, and Jeffrey P. Bigham. 2013. Visual challenges in the everyday lives of blind people. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '13), 2117–2126. https://doi.org/10.1145/2470654.2481291

[5] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2: 77–101. https://doi.org/10.1191/1478088706qp063oa

[6] Michele A. Burton, Erin Brady, Robin Brewer, Callie Neylan, Jeffrey P. Bigham, and Amy Hurst. 2012. Crowdsourcing subjective fashion advice using VizWiz: challenges and opportunities. In *Proceedings of the ACM SIGACCESS Conference on Computers and accessibility* (ASSETS '12), 135–142. https://doi.org/10.1145/2384916.2384941

[7] Denae Ford, Mahnaz Behroozi, Alexander Serebrenik, and Chris Parnin. 2019. Beyond the Code Itself: How Programmers Really Look at Pull Requests. In *IEEE/ACM International Conference on Software Engineering: Software Engineering in Society (ICSE-SEIS)*, 51–60. https://doi.org/10.1109/ICSE-SEIS.2019.00014

[8] Cole Gleason, Amy Pavel, Xingyu Liu, Patrick Carrington, Lydia B. Chilton, and Jeffrey P. Bigham. 2019. Making Memes Accessible. In *Proceedings of the ACM SIGACCESS Conference on Computers and Accessibility* (ASSETS '19), 367–376. https://doi.org/10.1145/3308561.3353792

[9] Cole Gleason, Amy Pavel, Emma McCamey, Christina Low, Patrick Carrington, Kris M. Kitani, and Jeffrey P. Bigham. 2020. Twitter A11y: A Browser Extension to Make Twitter Images Accessible. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 1–12. Retrieved January 12, 2022 from https://doi.org/10.1145/3313831.3376728

[10] Erving Goffman. 1959. The Presentation of Self in Everyday Life. Anchor.

[11] Jonathan Grudin. 1988. Why CSCW applications fail: Problems in the design and evaluationof organizational interfaces. In *Proceedings of the ACM Conference on Computer-supported cooperative work* (CSCW '88), 85–93. https://doi.org/10.1145/62266.62273

[12] Darren Guinness, Edward Cutrell, and Meredith Ringel Morris. 2018. Caption Crawler: Enabling Reusable Alternative Text Descriptions using Reverse Image Search. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 1–11. Retrieved January 16, 2022 from https://doi.org/10.1145/3173574.3174092

[13] Jeffrey T. Hancock and Catalina L. Toma. 2009. Putting Your Best Face Forward: The Accuracy of Online Dating Photographs. *Journal of Communication* 59, 2: 367–386. https://doi.org/10.1111/j.1460-2466.2009.01420.x

[14] Margot Hanley, Solon Barocas, Karen Levy, Shiri Azenkot, and Helen Nissenbaum. 2021. Computer Vision and Conflicting Values: Describing People with Automated Alt Text. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* (AIES '21), 543–554. https://doi.org/10.1145/3461702.3462620

[15] Noelle J. Hum, Perrin E. Chamberlin, Brittany L. Hambright, Anne C. Portwood, Amanda C. Schat, and Jennifer L. Bevan. 2011. A picture is worth a thousand words: A content analysis of Facebook profile photographs. *Computers in Human Behavior* 27, 5: 1828–1833. https://doi.org/10.1016/j.chb.2011.04.003

[16] Ju Yeon Jung, Tom Steinberger, Junbeom Kim, and Mark S. Ackerman. 2022. "So What? What's That to Do With Me?" Expectations of People With Visual Impairments for Image Descriptions in Their Personal Photo Activities. In *Designing Interactive Systems Conference* (DIS '22), 1893–1906. https://doi.org/10.1145/3532106.3533522

[17] Leqi Liu, Daniel Preotiuc-Pietro, Zahra Riahi Samani, Mohsen E. Moghaddam, and Lyle Ungar. 2016. Analyzing Personality through Social Media Profile Picture Choice. *Proceedings of the International AAAI Conference on Web and Social Media* 10, 1: 211–220. https://doi.org/10.1609/icwsm.v10i1.14738

[18] Kelly Mack, Edward Cutrell, Bongshin Lee, and Meredith Ringel Morris. 2021. Designing Tools for High-Quality Alt Text Authoring. In *Proceedings of the ACM SIGACCESS Conference on Computers and Accessibility*, 1–14. https://doi.org/10.1145/3441852.3471207

[19] 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 ACM Conference on Human Factors in Computing Systems* (CHI '17), 5988–5999. https://doi.org/10.1145/3025453. 3025814

[20] Valerie S. Morash, Yue-Ting Siu, Joshua A. Miele, Lucia Hasty, and Steven Landau. 2015. Guiding Novice Web Workers in Making Image Descriptions Using Templates. *ACM Transactions on Accessible Computing* 7, 4: 12:1-12:21. https://doi.org/10.1145/2764916

[21] Meredith Ringel Morris, Jazette Johnson, Cynthia L. Bennett, and Edward Cutrell. 2018. Rich Representations of Visual Content for Screen Reader Users. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 1–11. Retrieved January 15, 2022 from https://doi.org/10.1145/3173574.3173633

[22] Meredith Ringel Morris, Annuska Zolyomi, Catherine Yao, Sina Bahram, Jeffrey P. Bigham, and Shaun K. Kane. 2016. "With most of it being pictures now, I rarely use it": Understanding Twitter's evolving accessibility to blind users. In *Proceedings of the ACM Conference on Human Factors in Computing Systems* (CHI '16), 5506–5516. https://doi.org/10.1145/2858036.2858116

[23] Helen Petrie, Chandra Harrison, and Sundeep Dev. 2005. Describing images on the web: A survey of current practice and prospects for the future. In *Proceedings of Human Computer Interaction International (HCII)*, 71.

[24] Michael V. Reiss and Milena Tsvetkova. 2020. Perceiving education from Facebook profile pictures. *New Media & Society* 22, 3: 550–570. https://doi.org/10.1177/1461444819868678

[25] Elliot Salisbury, Ece Kamar, and Meredith Morris. 2017. Toward Scalable Social Alt Text: Conversational Crowdsourcing as a Tool for Refining Vision-to-Language Technology for the Blind. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing* 5: 147–156.

[26] Abigale J. Stangl, Esha Kothari, Suyog D. Jain, Tom Yeh, Kristen Grauman, and Danna Gurari. 2018. BrowseWithMe: An Online Clothes Shopping Assistant for People with Visual Impairments. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility* (ASSETS '18), 107–118. https://doi.org/10.1145/3234695.3236337

[27] Abigale Stangl, Meredith Ringel Morris, and Danna Gurari. 2020. "Person, Shoes, Tree. Is the Person Naked?" What People with Vision Impairments Want in Image Descriptions. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 1–13. https://doi.org/10.1145/3313831.3376404

[28] Abigale Stangl, Nitin Verma, Kenneth R. Fleischmann, Meredith Ringel Morris, and Danna Gurari. 2021. Going Beyond One-Size-Fits-All Image Descriptions to Satisfy the Information Wants of People Who are Blind or Have Low Vision. In *Proceedings of the ACM SIGACCESS Conference on Computers and Accessibility* (ASSETS '21), 1–15. https://doi.org/10.1145/3441852.3471233

[29] TJ Thomson and Keith Greenwood. 2020. Profile Pictures Across Platforms: How Identity Visually Manifests Itself among Social Media Accounts. *Handbook of Visual Communication.* 349–363.

[30] Violeta Voykinska, Shiri Azenkot, Shaomei Wu, and Gilly Leshed. 2016. How Blind People Interact with Visual Content on Social Networking Services. In *Proceedings of the ACM Conference on Computer-Supported Cooperative Work & Social Computing* (CSCW '16), 1584–1595. https://doi.org/10.1145/2818048. 2820013

[31] Jayne Wallace, John McCarthy, Peter C. Wright, and Patrick Olivier. 2013. Making design probes work. In *Proceedings of the ACM Conference on Human Factors in Computing Systems* (CHI '13), 3441–3450. https://doi.org/10.1145/2470654.2466473

[32] Shaomei Wu, Jeffrey Wieland, Omid Farivar, and Julie Schiller. 2017. Automatic Alt-text: Computer-generated Image Descriptions for Blind Users on a Social Network Service. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (CSCW '17), 1180–1192. https://doi.org/10.1145/2998181.2998364

[33] Lotus Zhang, Lucy Jiang, Nicole Washington, Augustina Ao Liu, Jingyao Shao, Adam Fourney, Meredith Ringel Morris, and Leah Findlater. 2021. Social Media through Voice: Synthesized Voice Qualities and Self-presentation. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1: 161:1-161:21. https://doi.org/10.1145/3449235

[34] Chen Zhao and Gonglue Jiang. 2011. Cultural differences on visual self-presentation through social networking site profile images. In *Proceedings of the ACM Conference on Human Factors in Computing Systems* (CHI '11), 1129–1132. https://doi.org/10.1145/1978942.1979110

[35] Yuhang Zhao, Shaomei Wu, Lindsay Reynolds, and Shiri Azenkot. 2017. The Effect of Computer-Generated Descriptions on Photo-Sharing Experiences of People with Visual Impairments. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW: 121:1-121:22. https://doi.org/10.1145/3134756

[36] How to Meet WCAG (Quickref Reference). Retrieved from https://www.w3.org/WAI/WCAG21/quickref/?versions$=$2.0#text-alternatives

[37] Accessible Digital Media Guidelines – Guideline A: Images. *Accessible Digital Media Guidelines.* Retrieved from https://www.wgbh.org/foundation/ncam/guidelines/accessible-digital-media-guidelines-guideline-a-images

[38] Seeing AI App from Microsoft. Retrieved from https://www.microsoft.com/en-us/ai/seeing-ai