Designing Accessible Obfuscation Support for Blind Individuals' Visual Privacy Management

Lotus Zhang
University of Washington
Seattle, USA
hanziz@uw.edu

Abigale Stangl
University of Washington
Seattle, USA
astangl@uw.edu

Tanusree Sharma
University of Illinois
Urbana, USA
tsharma6@illinois.edu

Yu-Yun Tseng
University of Colorado
Boulder, USA
Yu-Yun.Tseng@colorado.edu

Inan Xu
University of California at Santa Cruz
Santa Cruz, California, USA
inxu@ucsc.edu

Danna Gurari
University of Colorado
Boulder, USA
danna.gurari@colorado.edu

Yang Wang
University of Illinois
Urbana, USA
yvw@illinois.edu

Leah Findalter
University of Washington
Seattle, USA
leahkf@uw.edu

ABSTRACT

Blind individuals commonly share photos in everyday life. Despite substantial interest from the blind community in being able to independently obfuscate private information in photos, existing tools are designed without their inputs. In this study, we prototyped a preliminary screen reader-accessible obfuscation interface to probe for feedback and design insights. We implemented a version of the prototype through off-the-shelf AI models (e.g., SAM, BLIP2, ChatGPT) and a Wizard-of-Oz version that provides human-authored guidance. Through a user study with 12 blind participants who obfuscated diverse private photos using the prototype, we uncovered how they understood and approached visual private content manipulation, how they reacted to frictions such as inaccuracy with existing AI models and cognitive load, and how they envisioned such tools to be better designed to support their needs (e.g., guidelines for describing visual obfuscation effects, co-creative interaction design that respects blind users' agency).

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in accessibility**; **Empirical studies in HCI**; **Accessibility design and evaluation methods**.

KEYWORDS

accessibility, privacy-preservation technology, blind photography

ACM Reference Format:
Lotus Zhang, Abigale Stangl, Tanusree Sharma, Yu-Yun Tseng, Inan Xu, Danna Gurari, Yang Wang, and Leah Findalter. 2024. Designing Accessible Obfuscation Support for Blind Individuals' Visual Privacy Management. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24), May 11–16, 2024, Honolulu, HI, USA.* ACM, New York, NY, USA, 19 pages. https://doi.org/10.1145/3613904.3642713

CHI '24, May 11–16, 2024, Honolulu, HI, USA

ACM ISBN 979-8-4007-0330-0/24/05
https://doi.org/10.1145/3613904.3642713

1 INTRODUCTION

Photo sharing is an important activity for blind individuals to access visual information in daily life, socialize, keep memories, and express themselves [3, 13, 25, 28, 33, 81]. At the same time, visual privacy is a major concern with photos taken by blind people, due to accessibility challenges in reviewing and evaluating photos that contain private objects or information [3, 13, 25, 28, 54]. Blind individuals are thus exposed to higher security risks and more impression management issues when engaging in photo sharing compared to sighted counterparts [4, 5, 26, 81].

Recent research has begun to explore privacy protection features for blind people to manage their visual content (e.g., [8, 71, 87]). One promising approach is to leverage computer vision for detecting and obfuscating (e.g., blurring, removing) private content in blind individuals' photos [8, 16]. Accessibility researchers have interviewed blind individuals' perspectives on the use and design of this type of AI-assisted visual privacy tool, noting both needs and concerns from the community [8, 71]. In particular, blind people desire control over obfuscation decisions and would like tools to be designed more accessibly to support such control [8, 71]. Such a tool should inform blind users of potential private content in their photos and empower them to decide, manipulate, and evaluate the obfuscation of this content. Still, *what interaction designs could support these goals* and *how blind people would react to using such a tool* are under-explored questions.

To bridge this research gap, we prototyped and evaluated a mid-fidelity screen reader-accessible obfuscation tool, building on insights from past work (e.g., [8, 38, 71, 87]). To examine the capability of existing AI technologies in providing accessible visual obfuscation support while also exploring what an ideal system could offer, we prepared two versions of the prototype, one using off-the-shelf

AI models (i.e., Segment Anything Model [36], BLIP2 [42], ChatGPT [52]) and one using Wizard-of-Oz, human-authored guidance. We employed the prototype as a probe in a study with 12 blind participants. Participants edited private photos, including those from the BIV-PRIV [66] dataset (photos with fake "prop" private objects taken by blind people) and participants' own non-private photos. We focused on the following research questions:

- RQ1: When given the opportunity to apply computer vision methods for managing private visual content in photos, what are blind people's mental models of these methods?
- RQ2: How do blind people approach these computer vision methods and why?
- RQ3: What design opportunities exist to reduce friction in blind people's experiences with these methods?

Our findings reveal that blind participants had varied levels of pre-existing understanding of relevant visual concepts (e.g., background versus foreground, blur, inpainting) but were quick to learn and make use of these options. Still, participants experienced a range of frictions in using the prototype to manage private visual content (e.g., inaccuracies with the off-the-shelf prototype version, general difficulties with envisioning and evaluating obfuscation results, heavy cognitive load). Accordingly, participants offered design ideas to alleviate these frictions, such as allowing users to more freely make obfuscation decisions with only supporting input from AI and non-visual communications that clearly indicate obfuscated private content' visibility. We discuss how this feedback could inform both the design of more accessible visual obfuscation interfaces and adaptations to the underlying computer vision models that support blind users' sense-making of obfuscation effects.

In summary, our work makes the following contributions: (1) an empirical understanding of how blind people approach AI-assisted obfuscation manipulations for managing private visual content, (2) design insights for reducing frictions noted in blind people's use of AI-assisted visual privacy obfuscation tools, and (3) an example prototype design of an AI-assisted screen-reader accessible visual obfuscation tool.

2 BACKGROUND

Our research is informed by prior literature on blind photography, visual privacy, privacy-preservation technology, and accessible visual content-sharing support.

2.1 Privacy Concerns Related to Blind People's Photos

Blind people take and share photos for a range of purposes, from visual information access to social interaction and self-expression [3, 13, 25, 28, 33, 81]. These photos commonly feature text, outdoor scenery, people, food, vehicles, crafts, plants, household items, and so on [3, 33]. In sharing photos, however, blind people face challenges with reviewing and evaluating the photo content compared to their sighted counterparts [3, 13, 25, 28, 54]. As a result, photos shared by blind people often unintentionally contain private information [4, 5, 26, 81].

Blind people are aware of privacy risks involved in photo sharing and often feel concerned when engaging with cameras [4, 5, 13, 33]. Concerns include both inadvertently disclosing their own information and breaching others' privacy (i.e., multiparty privacy conflicts) [6, 7, 73, 89]. Recent work has examined blind people's privacy concerns and risks related to visual content sharing, a concept termed *visual privacy* (e.g., [7, 31, 72, 73]). Private visual content categories that blind people are particularly concerned with include: *financial* (e.g., bank account details, credit cards), *medical* (e.g., medical documents, prescription pill bottles, pregnancy tests), *people* (especially naked bodies and faces), and *location or identification* (e.g., digital screens, letters, papers with addresses and names) [7, 26, 73]. Blind individuals are also concerned about photos that may negatively influence others' perceptions of them (i.e., impression management), such as unflattering or embarrassing shots or unorganized homes, and activities that may be misinterpreted as bad behaviors [7, 73].

A range of factors influence blind people's comfort with photo sharing. For example, they are generally more willing to share with close friends and family [7]. In particular, they often work with sighted friends and family for visual information access and management [6, 87]. In doing so, however, blind individuals worry about the lack of independence as well as the potential for compounded risk of sharing sensitive information with close social ties [31, 87]. In turn, some individuals have become accustomed to sharing private content with remote visual interpretation services to access important information (e.g., Be My Eyes [2], Aira [1]), though their willingness to do so depends on the type of service as well as data handling and access policies [7, 71, 73]. Finally, other sharing considerations include: (1) the potential for disclosing bystanders' private information [7], (2) the impact on intimate personal relationships or broader social interactions if private information is disclosed [73], (3) the burden of choosing between the right to information access vs. others' profiting from their data [72], and (4) whether the information is shared knowingly (e.g., with a visual interpretation service to gain access to visual information) or inadvertently (e.g., in the background of a photo) [73]. These concerns mostly align with parallel visual privacy research efforts involving sighted people (e.g., [45]).

In this paper, we aim to advance technology design that gives blind people more control in managing their visual private content themselves through non-visual information access and photo manipulation. Additionally, blind people's visual privacy is generally researched within the context of visual interpretation services, yet their photo-sharing practices span a much wider range of contexts, many of which have been considered important by general privacy research (e.g., social media [12, 24]). Our research helps reduce this research gap by considering a range of common photo-sharing scenarios of blind people.

2.2 Accessible Obfuscation Design

Prior work has proposed a range of privacy-enhancing approaches, including but not limited to access control mechanisms (e.g., [23, 69, 79, 80]), privacy policy measures (e.g., [22, 78]), and privacy features that detect, flag, and limit sensitive information (e.g., [18, 44, 65]). Among them, *obfuscation* has been highlighted as particularly promising for protecting blind people's visual privacy [7, 8, 18, 26, 71]. Obfuscation is "the deliberate addition of ambiguous, confusing,

or misleading information to interfere with surveillance and data collection" [16], which allows hiding specific private areas in a photo while still displaying important information. Blind people also envision that obfuscation could help focus recipients' visual attention, which is useful in interactions with visual interpretation services [8, 71].

While obfuscation could be applied in many forms—such as blurring (e.g., [32, 67, 75, 84]), overlaying with stickers (e.g., [14, 43, 61]), silhouette/blacking out (e.g., [14, 53]), inpainting [51, 55], pixelating [29]—each form has pros and cons. For example, in an interview study, some blind people were concerned that blurring may be less reliable than blacking-out due to being more easily reversible, while some also value the potential for blurring to allow at least a vague understanding of the overall visual content without disclosing specific private information [71]. Obfuscation can also be applied to specific AI-detected private objects (e.g., [9, 39, 76, 85]), though such an approach is difficult, as privacy is contextual [10, 50]—so are the types of objects to be obfuscated [10]. Work with sighted users shows that people consider a utility-privacy trade-off when making obfuscation decisions about a photo [10, 29, 35], which has led to proposals to protect privacy without sacrificing utility, such as using avatars [46] and activity-oriented partial obfuscation [10].

Most existing obfuscation design work is geared toward sighted people, with only a recent focus on the needs of blind users. In an interview study, Alharbi et al. [8] explored blind people's perspectives on accessible obfuscation tool design. Their participants viewed obfuscation to be potentially useful but desired information and control over the obfuscation decision to ensure alignment between their intention and the automated obfuscation results [8]. Similarly, Stangl et al. [71] interviewed blind individuals on their expectations for obfuscation tools, noting hypothetical concerns around accuracy, processing delays, and reduced agency and control over their visual content. Moving beyond interview studies, Zhang et al. [87] developed ImageAlly, a prototype system that automatically detects private objects in an image, surfaces that information to a blind user, and, if desired, supports the user in handing off the image to a trusted ally for obfuscation, rather than allowing the blind user to edit the image independently.

Despite growing interest in obfuscation, prior work on supporting blind users in independently managing private visual content has been limited to interviews to capture projected perceptions. Our study instead explored how blind participants make use of an interactive prototype designed to support them in independently making decisions about when and how to obfuscate an image.

2.3 Accessible Visual Content Sharing Support

In the context of visual content sharing and editing, blind people tend to desire image descriptions beyond those typically recommended by general guidelines (e.g., [17, 57, 83]) and have more concerns around description accuracy [34, 63, 88]. For example, aesthetics and potential experiences triggered by photo content are considered relevant to photo-sharing decisions by blind individuals [30, 34, 63, 88]. Information related to spatial positions of objects and modifications on image content is critical for visual layout editing tasks [11, 56, 63]. Because of the abundance of visual details needed, cognitive load is a key challenge. Prior work suggested providing a quick, intuitive visual summary and opportunities to further explore image details [38, 63]. As many blind people lack understanding of visual concepts and design standards, support for learning in these areas is also important [41, 59]. Building on these insights, our study explored how non-visual image editing support should be designed to improve private photo obfuscation accessibility.

3 METHOD

We conducted 12 user studies to understand how blind individuals approach using computer vision methods to independently manipulate visual privacy obfuscation and gain actionable insights for accessible tool design. Participants used two versions of a mid-fidelity obfuscation prototype to manipulate private photo content: (a) a Wizard-of-Oz version (with functionalities pre-configured by researchers) for probing design feedback without distraction from algorithmic inaccuracies, and (b) an off-the-shelf version (implemented with newest off-the-shelf models) to explore how inaccuracies in AI models may influence participants' use of them. In this section, we describe the prototype and study design.

3.1 Prototype Design and Implementation

3.1.1 Prototype Design. Informed by blind users' desire to control obfuscation decisions with manageable cognitive load [8, 71, 87], this prototype provides support for users to non-visually manipulate private objects in photos. The prototype automatically detects user-specified private objects in photos and allows users to decide whether and how to obfuscate them. The prototype is presented through a simple one-page user interface that prioritizes easy navigation using screen readers and consists of two main components: (1) an *explore image* section and (2) an *edit image* section (as shown in Figure 1).

Explore Image: The prototype first presents users with a high-level caption and then a touch-based explorer for learning photo layout—as users touch different areas of the photo, object names and text surrounding the area are announced. We designed this feature to provide photo descriptions at different granularity in supporting better interpretation, as inspired by [38].

Edit Image: The prototype detects potential private objects and displays each object's caption in the 'Item' drop-down menu (without a particular order). From this menu, users could choose to obfuscate any detected objects or the background of the photo focus. Users could also configure a small set of most common obfuscation settings (based on [9, 29, 44]): from the 'Style' drop-down menu, users can choose to blur, blackout, or erase (i.e., removing a private object and inpainting background to fill the area) the private content; from the 'Shape' drop-down menu, they can configure the obfuscation area either to fit the exact shape of the private object or to be a bounding box (rectangle) fully enclosing the object. We use this set of options as a starting point to elicit participants' preferences for visual obfuscation manipulation choices. Upon applying the obfuscation, users can review the resulting photo in the explore image section.

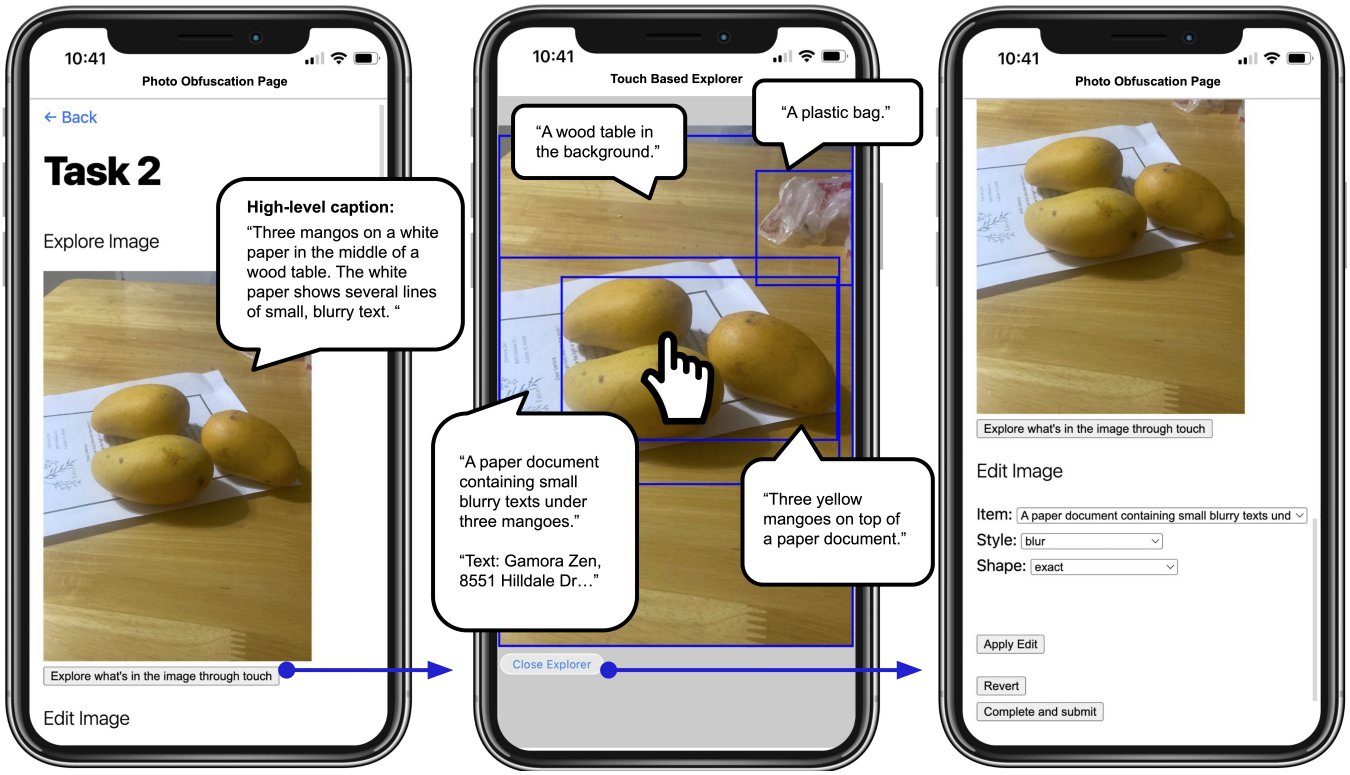


Figure 1: Prototype user interface design for (1) explore image section, including a high-level caption (left) and a touch-based explorer with captions displayed for each object bounding box (middle), (2) edit image section (right). The prototype allows users to explore the image and obfuscate private objects by choosing from a list of all potential private objects detected by the system. The interface was designed to mimic a standalone smartphone app but was implemented as a webpage for study participants' easy access. Displayed captions are from the Wizard-of-Oz version. Please see Supplementary Materials for object captions from the off-the-shelf version.

3.1.2 Prototype Implementation. We implemented two versions of the prototype that differed in underlying methods to describe image content and detect private objects:

Off-the-shelf Version: This implementation employed existing AI models for describing photos, identifying private objects, and obfuscating those objects. This version allowed us to understand whether and how existing AI models can support non-visual obfuscation interpretation and decision-making, with a focus on blind participants' reactions to likely inaccuracies. We adopted an approach used by the Caption-Anything image processing tool [82] by employing the Segment Anything Model (SAM) [36] and BLIP2 [42] to locate and caption objects in an image. We first used SAM to segment all objects in an image, constructed a bounding box for each segment, and cropped the image area within the bounding box to feed into BLIP2 for caption generation. Then, to detect whether an object belongs to any user-specified private categories, we used the ChatGPT-3.5 API [52] to process each caption with the prompt: "Does the following sentence mention anything related to a [private object category]? Answer yes or no. The sentence is: there is [caption]." This implementation choice was informed by two considerations: (1) SAM and BLIP2 both do not require training on blind people's private images (the collection of which can be challenging and potentially unethical) and thus produce better results on photos used in our study than common off-the-shelf object detection models; (2) SAM and BLIP2 are easily accessible and do not require additional fine-tuning as most state-of-the-art models do, which increases the replicability of our implementation approach for future research. For consistency, our prototype generated high-level captions through BLIP2 as well.

Although this processing pipeline is capable of detecting any private object category, we limited our implementation to five private categories that previous work identified as especially concerning for blind individuals: (1) medical, (2) financial, (3) personally identifiable information, (4) impression management-related, and (5) faces [7, 26, 73]. Limiting our set to five object categories was useful for two practical reasons: (1) by instructing all participants to focus on obfuscating the same private object categories, we could more consistently analyze their reactions to the prototype design; (2) models for segmenting and captioning images often require extensive computational power and can cause unreliability during user studies—limiting private object categories allowed us to preprocess images prior to the study sessions. To achieve improved model performance, we used concrete object names to represent the five abstract private object categories in the ChatGPT prompts:

credit card, pill bottle, human, sexual product, and paper document. We intentionally left the paper document a broad category to allow users to decide the privacy risks of the document content themselves, as suggested by [8].

The off-the-shelf models used in this prototype version tend to produce the following inaccuracies: (a) BLIP2 could mis-categorize objects, such as "mangoes" as "potatoes"; (b) SAM could segment sub-parts of an object, leading to duplicated objects identified by the prototype (we used intersection-over-union to remove duplicates but left smaller sub-parts in case users want to hide only a small area of an object); (c) BLIP2 could inaccurately describe unclear or abstract image areas, such as those with obfuscation effects (e.g., blurred). We focused on understanding how participants react to these inaccuracies in the study.

Wizard-of-Oz Version: To understand what participants' experiences could be in the future with even more accurate underlying models, we implemented a Wizard-of-Oz version of the prototype. In this version, researchers assessed the results of off-the-shelf models and authored a ground truth for each result. Due to the high performance of existing tooling for optical character recognition, image visual effect application, and image segmentation, we limited the scope of the Wizard-of-Oz components to two automation tasks: (a) identifying, locating, and describing objects and (b) detecting private objects. As the off-the-shelf version, these researcher-annotated results (e.g., all objects' descriptions and bounding boxes, detected private object list) were manually inserted into the prototype system prior to the study, so that participants could operate the two prototype versions in the same manner. Two researchers collaboratively generated this ground truth information. Following image description best practices (as suggested by [17, 27, 57, 74, 83]), we decided on six object description rules: (1) for each object, focus on describing what the object is, its salient characteristics (e.g., color, identity, number, pattern), and actions; (2) if the object's bounding box includes another object underneath the primary object, describe this spatial relationship (e.g., a black cat lying on the couch); (3) if there are multiple objects of the same type close to each other, provide one description for them to avoid confusion (e.g., one description for three mangoes, instead of three descriptions for each mango); (4) for obfuscated areas, describe their corresponding visual effects, shapes, and colors if relevant (e.g., a blurry rectangle with yellow and white colors, a black human silhouette); (5) for visual artifacts left from obfuscation (e.g., unnatural in-painting), briefly describe what the unnatural area looks like to the annotators (e.g., a moving, blurry blue object); (6) for high-level image caption, describe all salient objects in the image. These description rules served as a starting point for us to explore image description best practices in the context of visual obfuscation manipulation.

For both the off-the-shelf and Wizard-of-Oz version, we programmatically applied visual obfuscation effects—(1) blackout: set all pixels of the obfuscation area black; (2) blur: applied a Gaussian blur with a high radius value (80) to the obfuscation area; (3) erase: inpainted the obfuscation area with surrounding background using the LaMa (large mask method) tool powered by the SOTA AI Model [62, 77]. The touch-based explorer additionally featured Microsoft Azure AI's optical character recognition model [48] for text detection. To ease access for study participants, we implemented this mobile application prototype as a webpage and instructed participants to use it through their smartphones. A demo video is included for this prototype in the Supplementary Materials.

3.2 Participants

We recruited 12 blind participants through the National Federation of the Blind mailing list and word of mouth. Participants had to be at least 18 years old, identify as blind or legally blind, and have experience taking photos. To ensure consistent screen reader behavior with our prototype, we limited recruitment to iPhone users. As shown in Table 1, participants were 24–59 years old ($Median = 36$, $M = 40.2$), with eight identifying as female and four as male (open-ended description for gender), and all self-reporting to be either totally blind (N = 7) or with some light perception (N = 5). Their visual condition onset ranged from birth to 40 years old, with the majority beginning at birth (N = 7). Five participants had no visual memory, three had limited visual memory, and four had significant visual memory. In terms of photo-sharing experiences, participants most commonly shared photos 'once a week' (N = 7), followed by 'once a month' and 'less than once a month' (N = 2 for both), with only one (P7) sharing once a day or more. Participants' experience with photo editing was more limited—the majority had never edited a photo (N = 8), with three editing less than once a month and one editing approximately once every month. For photo sharing and editing, participants used mobile phones (N = 12) and desktop or laptop computers (N = 3) with screen readers (N = 11) as well as remote sighted assistance (N = 8).

3.3 Study Protocol

Participants filled out a short pre-study survey about their demographics and experiences with photography tasks, and then participated in a 90-minute remote study session via Zoom. The full protocol is included in the Supplementary Materials. Below, we detail the photo choices for the obfuscation tasks before summarizing the study procedure.

3.3.1 Obfuscation Task Photo Selection. To provide a degree of ecological validity for the study tasks, we selected photos primarily from the BIV-PRIV dataset [66]—a dataset that contains photos taken by blind people of fake "prop" private objects, such as medical and financial documents, pill bottles, and sensitive objects that could raise privacy concerns (e.g., condoms, pregnancy tests).

We selected one photo from each of five especially concerning categories (i.e., (1) medical, (2) financial, (3) personally identifiable information, (4) impression management-related, and (5) faces). The first four of these photos were from BIV-PRIV, while the face photo is a stock photo from [58], as photos with faces were not included in BIV-PRIV. We used one of these five photos (a credit card) to *familiarize* participants with the prototype (Table 2), while the remaining four photos were reserved for the *main photo obfuscation tasks*. Table 3 shows the image descriptions and object detections for each of the main obfuscation task photos for both the off-the-shelf and Wizard-of-Oz prototype versions. Each of these photos was presented to participants alongside a photo sharing scenario, such as *"You took a photo of some newly bought mangoes for a fruit review post on social media"* for photo (c) and *"You took a photo of your new office space to post on social media"* for photo (a) (Table 3).

Participant	Gender	Age	Visual Condition	Onset	Visual Memory?
P1	Female	35	Totally Blind	Birth	No
P2	Female	37	Totally blind	Birth	No
P3	Female	28	Some Light Perception	4 years old	Limited
P4	Female	34	Totally Blind	Birth	No
P5	Female	44	Totally Blind	Birth	No
P6	Male	29	Some Light Perception	Birth	Limited
P7	Female	24	Totally Blind	6 months old	Limited
P8	Female	55	Totally Blind	18 years old	Yes
P9	Male	28	Some Light Perception	Birth	No
P10	Male	55	Some Light Perception	40 years old	Yes
P11	Female	54	Totally Blind	Birth	Limited
P12	Male	59	Some Light Perception	13 years old	Yes

Table 1: Participants' demographic information.

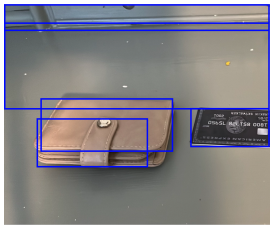
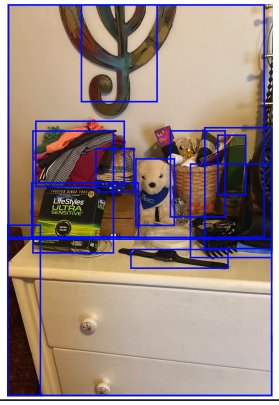
Familiarization Task Photo		Detection Error Demo Photo	
	**Photo caption:** a credit card and a wallet sitting on a table **Private object:** the jpmorgan chase black card (credit card)		**Photo caption:** a white dress with a stuffed bear on it **Private object:** none

Table 2: Photos used in familiarization task (left, showing a credit card) and to demonstrate a detection error (right, showing a condom box that is undetected as a private object). For these two photos, we only presented caption and private object detection results (associated privacy category indicated in the bracket) from the off-the-shelf implementation to participants, as the left photo was used only for introducing participants to the interface elements and the right photo was meant to show potential off-the-shelf model inaccuracies.

We also selected one *detection error demo photo*. While the off-the-shelf models made minor mistakes in describing most photos from the BIV-PRIV dataset, they produced more false negative object detections for photos that were visually crowded. We included one of these photos where the prototype missed detecting a condom box (Table 2) to understand how participants react to this type of inaccuracy.

Last, we included a *personal non-private photo*. In the pre-study survey, participants had an option to voluntarily upload a photo they had taken recently to edit with our prototype. To protect participants' privacy, we asked for this photo to not contain any actual private information, but instead, the researcher picked a non-private object from each photo's background to ask participants to obfuscate. While this approach did not provide an opportunity for participants to obfuscate their private information, it allowed us to learn about how participants may experience our prototype differently with their own photos compared to others'.

3.3.2 Procedure. The study session was conducted via Zoom and included three parts: (1) initial understanding and familiarization task, (2) main image obfuscation tasks, and (3) post-study interview. Participants were required to join the Zoom call from a smartphone and share their phone screen during the study tasks (with consent). Prior to the study, we emailed them instructions for accessing the mobile prototype website and asked them to keep the site open in a browser tab when joining the call to avoid additional browsing that may increase privacy disclosure risks. With a researcher's support, all participants were successful in setting up the study environment.

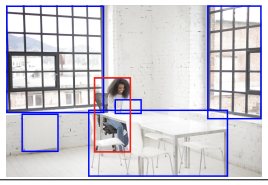
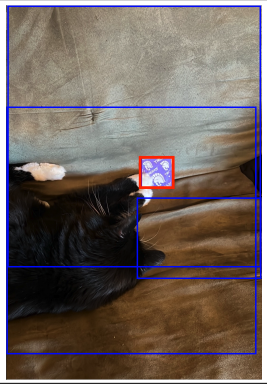
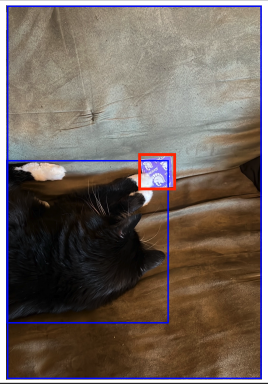
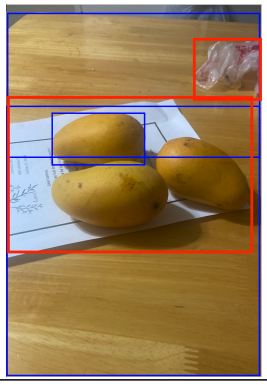
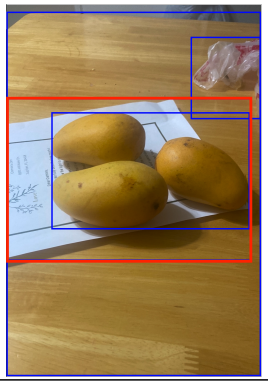
	Off-the-shelf Result		Wizard-of-oz Result	
(a)		**Photo caption:** a woman sitting at a table with a laptop **Private object:** (1) a woman sitting at a table with a laptop (human); (2) a woman sitting at a table with a laptop (human)		**Photo caption:** a room painted white, with two large windows and a woman sitting at a table surrounded by four chairs **Private object:** a smiling women sitting next to a table, working on a laptop (human)
(b)		**Photo caption:** a couch with a blue and black floral pattern **Private object:** a blue bottle of pills sitting on a bed (pill bottle)		**Photo caption:** a close up of a beige colored couch in floral patterns, with a blue pill bottle in the corner. The pill bottle has blurry small texts. **Private object:** a blue pill bottle with small blurry texts on the couch (pill bottle)
(c)		**Photo caption:** a cat laying on a couch **Private object:** a person holding a purple and white towel with a spartan logo on it (human)		**Photo caption:** A black cat lying on a brown colored couch, with a plastic condom bag near its paw. **Private object:** a purple plastic condom bag (sexual product)
(d)		**Photo caption:** three mangoes sitting on a piece of paper **Private object:** (1) three mangoes on a piece of paper (paper document) (2) a plastic bag with a dog inside (human)		**Photo caption:** three mangoes on a white paper in the middle of a wood table. The white paper shows several lines of small, blurry texts. **Private object:** a paper document containing small blurry texts under three mangoes (paper document)

Table 3: Overview of the four photos used for the main photo obfuscation tasks, showing descriptions and private object detection results (associated privacy category indicated in the bracket) from both the off-the-shelf and Wizard-of-Oz implementations. Each participant edited two of these four images using the prototype, in addition to the detection error demo photo (Table 2) and an optional personal photo that they could bring themselves. The off-the-shelf version of photo (a) shows an example of multiple bounding boxes inaccurately detected for one object.

Initial understanding and familiarization task: The researcher first guided participants through editing the credit card photo shown in Table 2 using the off-the-shelf prototype version. In introducing the prototype, we asked participants to imagine they had configured the system to detect the five private object categories (i.e., credit card, pill bottle, human, sexual product, and paper document) and explained that in the future it could be configured to detect other specific objects of interest.

To gauge initial understanding of relevant visual concepts, participants were instructed to read through all obfuscation options in the 'Edit Image' section and describe what they expected each option would do to the photo.

The researcher then provided a verbal description to clarify each option, along with a tactile metaphor for the obfuscation styles as follows:

- **Background vs. primary object:** *"...hiding everything behind the most prominent object in the image, if there is one. For example, if you took a photo of an apple on a kitchen counter-top, the apple should be the primary object, and everything else on the counter-top is considered background."*
- **Obfuscation style:**
  - **Black out:** *"Removing the content a user wants to hide, leaving the area black"*, with tactile metaphor, *"Imagine a plastic plate with tactile patterns that depict the shape of the United States. We cut a part of the plate so that you can't tell that the depicted shape is of the United States anymore, but you can feel a hole on the plate."*
  - **Blur:** *"Making content that a user wants to hide less clear by adding noise to the area of the image"*, with tactile metaphor, *"putting a soft fabric on top of the same plastic plate, so that you can feel the shape on the plate less clearly."*
  - **Erase:** *"Removing the content a user wants to hide, and filling it in with non-sensitive content that naturally blends into the photo,"* with tactile metaphor, *"we again cut a part of the plate but replace it with another piece of plastic with a different outline that blends into the rest of the plate seamlessly."*
- **Bounding box vs. exact shape:** *"Choose the hidden area to either exactly fit the shape of the private object, or a rectangle that encloses the object."*

After the familiarization session, researchers examined participants' understanding of these concepts again by asking them to provide a definition for each in their own languages.

Image obfuscation tasks: Participants independently reviewed and obfuscated three to four photos, depending on whether they opted to work on a personal photo. They were instructed to think aloud during the tasks and "make decisions about what you want to do in each scenario based on your feelings, judgment, and relevant past experiences—imagine you would share the obfuscated photo on social media." The first two photos were randomly assigned from the set of four main task photos (ensuring an equal number of participants to process each photo). At the end of editing each photo, we asked participants about: (a) considerations in deciding what/how to obfuscate (e.g., "How did you decide that this image task is completed?", "Why did you decide to manipulate the image this way?", "If you were sharing this image to a [coworker/visual interpretation assistant] instead, would you edit the image differently? How?"); (b) experiences with the obfuscation interaction (e.g., "How would you describe your experience of exploring and editing images with our system so far?", "How ready would you feel if you were to share the photo?"); (c) design feedback (e.g.,"How useful or not do you find the information provided for this photo?", "What additional information would you like to know? What suggestions do you have for presenting information?"). For the two main task photos, participants were given the off-the-shelf prototype version for one photo and the Wizard-of-Oz version for the other (order counterbalanced). They were initially generally informed that the two tasks made use of different algorithms and were given more information upon the completion of both tasks: "task _ is an ideal version of the tool that works fully accurately, whereas task _ is the version that is currently possible through existing algorithms, which can be inaccurate". They were then instructed to try out the off-the-shelf prototype with the detection error demo photo (Table 2) and share how they envision such inaccuracies to influence their use of AI-assisted privacy obfuscation tools. Last, participants who opted in also tried out the prototype (off-the-shelf version) on their own photos.

Post-study interview: Participants were asked how they felt about the overall idea of using this type of application to support their visual privacy management needs. The researcher also probed for benefits and frictions they foresee in using this type of tool and how participants may use it differently in real life. The interview ended with an open-ended conversation on how the system could be better designed.

3.4 Data and Analysis

Our data collection happened through (1) observational notes, (2) screen recordings of participants' interactions with the prototype, and (3) transcribed audio recordings of the study. Upon study completion, all video clips irrelevant to the prototype interaction were removed, and the remaining clips were cropped to contain only the prototype interface. We adopted a thematic analysis approach in analyzing our qualitative data, as outlined by Braun and Clark [15]. The first author reviewed all transcripts and observational notes to develop an initial codebook and coded through all data. The second and third author then randomly selected half of the coded transcripts to review. They then collaboratively iterated on the codebook and extracted key themes. The first and fifth author noted down all user actions in using the prototype to triangulate with the qualitative data.

4 FINDINGS

We report on participants' understanding of visual obfuscation concepts, workflow with the obfuscation tasks, aspects of the prototype they found to be challenging, as well as the related design feedback they provided.

4.1 Understanding of Visual Obfuscation Concepts (RQ1)

Pre-existing understanding: The majority of participants (N = 9) learned about the concept of hiding private content in photos

prior to attending the study sessions, though none of them had ever obfuscated content in a photo. Specifically, participants had different levels of pre-existing understanding of visual concepts involved in obfuscation manipulation. First, half (N = 6) of the participants clearly understood the difference between *foreground and background* (e.g., *"space like surrounding the main objects in the photo"* (P7)), while the other half felt vague or confused about it: *"I'm not sure what the whole background is, I guess...any furniture, people, or anything that are in the background?"* (P3). Among the obfuscation styles, *blacking out* was most commonly understood (N = 12), followed by *blurring* (N = 8), and lastly *erasing* (i.e., removing a private object and inpainting background to fill the area) (N = 4). Participants were particularly unsure about what would happen to an area once private objects were erased, for example: *"Does that get rid of it? Like maybe just take it out altogether...I don't know if there would be anything for you to see though"* (P5). Last, most participants (N = 8) also understood how obfuscation could be applied to different locations and differently shaped areas, though some were confused about the relative sizes of bounding boxes compared to the enclosed objects. Participants' pre-existing understanding partially came from personal visual memories (e.g., P11 considered blurring familiar because her previous vision as blurry) and conceptual knowledge, such as *"from kind of the concept and analogies"* (P6).

Understanding after explanation: With a verbal description and tactile metaphor, all participants felt generally confident in understanding the above-mentioned visual concepts and were ready to make obfuscation decisions accordingly: *"I feel like the descriptions are very easily understandable...I can grasp very quickly what I need to do"* (P4). However, they found envisioning some obfuscation results to be less straightforward, such as more complex combinations of options (e.g., the background of a bounding box) and the outcome of erasing. Participants generally understood the erasing option as being able to *"completely getting rid of"* (P2, P3, P6, P7, P10) the private object, but some could not envision what the area would look like after: *"Would there still be like the floral print in the background? Would it erase the pill bottle? And then there wouldn't just be this random spot on the couch?"* (P3). P7, for example, assumed *"a blank spot"* (P7) in the result photo. Confusions on these options sometimes caused participants to avoid selecting them.

In summary, our participants were comfortable learning about common visual concepts involved in obfuscation manipulations, while many had related pre-existing understanding. However, envisioning certain obfuscation results can be challenging, requiring the obfuscation tool to provide effective communication.

4.2 User Workflow and Decision Making (RQ2)

4.2.1 Workflow Overview. Figure 2 presents a summary of participants' general workflow. All obfuscation tasks began with an *initial exploration* of the original photo. On average, each initial photo exploration took 84.9 seconds (Min = 21; Max = 272; SD = 71.6). Typically, participants made use of a combination of the high-level caption and the touch-based exploration (N = 11). In most cases, participants quickly checked the high-level caption to get a general idea about the photo and used the explorer feature for further details. For some tasks, they also went back and forth between the two to understand how they corresponded to each other (N = 4). Occasionally, only one of the two features was used (explorer only for 5 out of 41 tasks; high-level caption only for 3 out of 41 tasks). P6, for example, felt that the high-level caption was enough for his initial exploration of photo 2: *"I didn't necessarily feel like I needed to know the placement of the images at the moment"* (P6).

In using the touch-based exploration, most participants wanted to gain an initial understanding of the photo layout (N = 10) and identify potential private objects (N = 9): *"just trying to see if there was anything potentially that could give off any information"* (P7). The majority tried gaining this information by both (1) touching different areas for spatial information and (2) swiping left and right to go through all objects—*"Swiping is definitely easier but...you wouldn't know exactly where it was"* (P8)—though many would like to directly hear verbal spatial descriptions. From this exploration, participants were sometimes able to quickly detect and locate private objects, but were also often left unsure, especially with incorrect or insufficient AI-generated captions from the off-the-shelf version of the prototype.

After exploring the original photo, participants then focused on whether and how to *apply obfuscation*. The total time spent making obfuscation decisions for one photo was on average 128.9 seconds (Min = 59; Max = 260; SD = 63.5). Participants relied on both their own judgement from the earlier exploration (e.g., *"based on the description, it looks like what I want is just to show the mangoes"* (P6)) and the system-detected private objects in forming the obfuscation decisions (detailed considerations in Section 4.2.3). After applying an obfuscation, participants always reviewed its effects through the high-level caption (14 out of 41 tasks) and/or touch-based exploration (all tasks). For approximately 60% of all tasks, participants were able to make the final decision in one try, whereas for the remaining 40%, they were less sure what manipulations would work best and adopted a trial and error approach by testing out and reviewing a range of options (N = 9), as P4 did to ensure that the focus of the photo, a cat, was unaffected by the obfuscation: *"I will black it out this time, and I'm gonna hide the exact shape just to see what it does, partly cause I don't want it to get on the cat."* The highest number of obfuscation adjustments our participants made for one photo was seven times by P5, followed by P9, who also changed and reviewed his obfuscation five times: *"I was experimenting with how each one will be"* (P9).

In reviewing obfuscation changes, participants in general needed to explore through all objects to evaluate the area that had been affected and ensure the absence of private information (N = 12). Some (N = 3) attempted to identify the exact change efficiently by memorizing the location of the private area in the original photo: *"I could tell kind of where I needed to touch to know that object had been hidden"* (P4). However, this approach was not always reliable and at times led to misunderstanding. For example, P8 missed a metal door near the private object when she initially explored an photo, so when she heard a door announced post-obfuscation, she thought it was an outcome of the obfuscation process: *"I think where the card was, it now describes it as a middle door with a hole in it...it's funny how it picks up different things"* (P8). P1 and P11 also tried to use the order of objects appearing in the image explorer as an anchor to track where the changes were supposed to happen, though our prototype was not designed to keep objects in the same order and

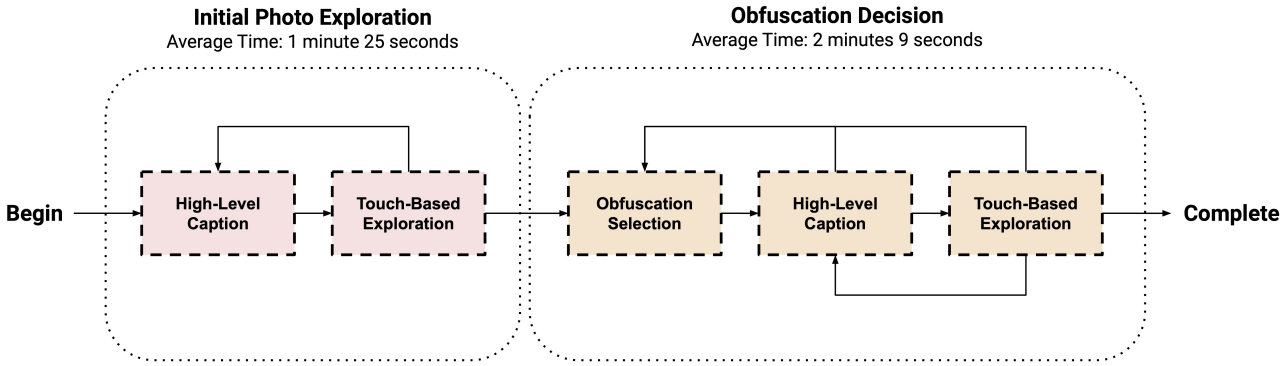


Figure 2: An overview of participants' general workflow in using the prototype to obfuscate private content in photos.

thus did not support this approach. Participants also used various strategies to check potential inaccuracies (N = 10), such as noting inconsistency in captions (N = 9) and using common sense: *"It could be a toy or something. I don't think you'd actually put a dog in the plastic bag"* (P9).

4.2.2 Experience with the Study Prototype. Overall, participants found the prototype easy to operate (N = 11). However, they did not feel confident enough that they would be willing to share all photos they had obfuscated. Only 11 obfuscation results from the 24 main photo tasks were considered ready to share on social media, with six of them deemed absolutely not shareable, and six of them deemed difficult to decide (one task was not completed due to a technical issue).

While participants found both the off-the-shelf and Wizard-of-Oz prototypes to be straightforward, many mentioned having a more positive experience with exploring and obfuscating photos in the Wizard-of-Oz tasks (N = 8)—they felt clearer about whether the result photos were shareable or not: *"When I tried to remove the bottle, it still described it as like a bottle being there (Table 5), so I knew you could still kind of see it"* (P3). In contrast, participants found it difficult to judge the result of an obfuscated photo when the off-the-shelf tools provided inaccurate descriptions: *"...it said something about there being a blank laptop screen...it didn't really make a lot of sense to me*" (P3). Section 4.3.1 provides further description of this specific concern. Still, even with the Wizard-of-Oz prototype version, participants mentioned a range of frictions that they experienced, including ineffective obfuscation communication (Section 4.3.2) and high cognitive load (Section 4.3.3).

4.2.3 Obfuscation Considerations. Table 4 presents the obfuscation choices made by participants across four private content types in the independent image obfuscation tasks. In terms of the obfuscation style, participants generally preferred *blacking out* or *erasing* but commonly chose *blurring* for human faces (5/5 participants who completed the task). For obfuscation shape, the *exact* shape seemed to be overall preferred, as chosen by all for both paper documents and sexual products, by all but one for human faces, and all but two for pill bottles. For the two tasks where participants had the option to choose between obfuscating the image foreground and background, more of them chose the former.

We identified four themes related to factors that affected participants' obfuscation decisions: (1) privacy, (2) information delivery, (3) visual presentation, and (4) context considerations.

First, all participants considered the level of perceived privacy an obfuscation edit provides for *different private content types*. For example, when dealing with highly sensitive private objects, such as a pill bottle or a condom package, participants commonly felt that erasing the object would be the safest, as blacking out and blurring both risk catching viewers' attention and suggesting the appearance of something private: *"I just figured it would draw less attention...your eyes would go to the blurry part to try and make out what it was"* (P7). Seven participants avoided using blurring, as viewers *"could sort of squint and see"* (P4) the hidden content and that technology could *"take that image and bring it more into focus to where it can be read"* (P11). Participants were also aware of risks related to disclosing the shapes of private objects. For example, P8 was concerned that the shape of a person alone may be identifiable and decided to use the bounding box option: *"to kinda hide a little more as to who the person was."* When participants were particularly concerned about privacy disclosure or unsure about private content location, they chose to hide the entire background: *"Recognizing that there's a lot of clutter with some text that may or may not be there, it was just easier to almost aggressively hide everything"* (P6). Participants' willingness to share the resulting images after obfuscation also varied across private content types. In particular, the majority of participants who worked on the photo (b) (couch with a pill bottle, as in Table 3) did not want to share it even after obfuscation (4 votes out of 6). Many mentioned that they were especially concerned about revealing private medical information, such as: *"I'm not as ready as I would like just because the bottle is still on the couch...it doesn't say that they can see the text, but I would still be kind of cautious because that was a lot of information"* (P5). In comparison, the other obfuscated photos all received more votes for being ready to share compared to not shareable.

Second, participants were commonly concerned about obfuscation affecting information a photo is supposed to deliver (N =

Private Object Category	Style			Shape		Item*	
	Blackout	Erase	Blur	Exact	Bounding Box	Foreground	Background
Pill Bottle	3/6	2/6	1/6	4/6	2/6	-	-
Paper Document*	3/5	1/5	1/5	5/5	0/5	3/5	2/5
Human*	5/5	0/5	0/5	4/5	1/5	-	-
Sexual Product*	2/5	2/5	1/5	5/5	0/5	4/5	1/5

Table 4: Participants' obfuscation choices across private content categories used in the independent image obfuscation tasks. Obfuscation choices are ranked in popularity within each category. In summary, blackout style, exact shape, and foreground seem to be the most popular, though the choices varied based on private content categories and related considerations, as described in Section 4.2.3. *For the two tasks where the image focus was the background (e.g., a whole office space), participants were not provided the option to obfuscate the entire background. *One participant (P8) chose not to make any edits for two image tasks and one participant (P12) did not complete one of the tasks due to a technical issue. Therefore, three of the four categories only had obfuscation choices from five participants, whereas the other one had six.

11). For example, P12 considered how hiding an entire credit card may impede the original photo sharing goal—to find the owner of a lost wallet: *"Not everybody is gonna have a Morgan Chase card. I don't. But I've got a tan wallet. When I can show someone a picture on social media of both items without identifying information that narrows down the population of people that it likely belongs to"* (P12). He therefore hoped to hide just the text on the card but not the card design. Some participants also wanted to hide more content to prevent irrelevant content from distracting the information they intend to deliver: *"there's a lot of stuff that people don't really need to know that I would almost want to hide everything but the animal"* (P5). A number of them (N = 4) associated certain obfuscation styles with cultural meanings that they considered appropriate for only specific scenarios: *"I know that they used to black out people's eyes in police lineups...that would at least show that there was something there but we can't tell you anything about it"* (P10).

Third, some participants considered the visual presentation of obfuscated photos in making decisions (N = 9). For example, participants were concerned about the resulting photo looking *"weird"* (P5, P6), *"funny"* (P5), or *"unnatural"* (P1, P2). Specifically, P1 and P2 considered blacking out the background behind a cat unnatural: *"I didn't want the cat to be shown in the air"* (P1). P9 was also concerned that mangoes on top of a piece of blacked out paper will appear as *"burned mangoes"* (P9). Participants occasionally wanted the obfuscated photos *"clean"* (P6, P8) and *"attractive"* (P3). For example, P6 mentioned wanting to *"edit that image even further and maybe put the mangoes in the center of the image,"* and P3 considered removing unattractive items, such as *"a radiator"* (P3) if the photo was meant for showing a new office space.

Participants mentioned that the above considerations would likely shift across photo-sharing contexts. For example, when the recipients are remote sighted assistants, participants generally felt less privacy concerns with sharing unaltered or blurred photos: *"cause I mean they signed confidentiality agreements"* (P4). In contrast, participants wanted safer obfuscation options (e.g., entire background, black out, erase) when sharing with coworkers and social media, depending on how close their relationships are: *"...if I'm sending the picture to the news, it's gotta be perfect. If I'm putting it on social media, it's gotta be close, and then if it's to send it to my brother...he's not gonna care if there's something in there"* (P10). Further, some participants only felt the need to refine their photos in more formal or public occasions—for example, P6 would only consider cleaning up a photo if it was for social media, and similarly, for business-related photos, P2 wanted to *"make sure that (the photo) looks professional before I post it."*

4.3 Frictions and Design Insights (RQ3)

Participants experienced a range of frictions in reviewing and obfuscating the photos, some innate to using AI-assisted photo obfuscation tools and some likely addressable through design.

4.3.1 Effects of Inaccuracy with the Off-the-shelf Prototype. Despite strategies for identifying inaccuracies (as described in Section 4.2.1), we observed that all participants were misled or confused by *captions* sometimes generated from the off-the-shelf models. For example, the models tend to generate inconsistent object captions for an object (e.g., mangoes described as lemons or potatoes, a stuffed animal as a dog or bear) when its surrounding area was obfuscated differently—with which, participants felt insecure: *"I can't be sure if it was a teddy bear or a dog... the first one told me it was a stuffed teddy bear and then the second one told me it was a dog"* (P12). For objects that were consistently mis-described (e.g., a purple condom bag as a purple toy), participants had no clue that there was inaccuracy: *"Well, obviously I got confused. Because I took it for its word"* (P12). Inaccurate captions led to the most confusion when describing an *obfuscated* part of a photo. Participants found it surprising and *"ridiculous"* (P10) that the caption model attempted to identify blurred, blacked-out, or distorted areas as objects, and in doing so producing false positive object detections (as Table 5 presents): *"It is kinda strange how it's different depending on which way you hide it...more things are appearing"* (P10). Participants were sometimes unsure what the system actually did and thus were hesitant to share the photo: *"It's hard to tell, you know, what is accurate and what isn't"* (P6).

For inaccuracies related to private objects being *mis-detected*, participants had mixed feelings. Some considered such errors unsurprising: *"I guess it would be tough to have it recognize all kinds of boxes for some stuff in it. So I don't think it [detection errors] can*

	Original	Erase & Exact	Blackout & Bounding Box	Blur & Exact
Off-the-shelf				
Image Caption	A couch with a blue and black floral pattern.	A person sitting on a couch with a laptop.	A close up of a couch with a floral pattern.	A white and black floral patterned couch.
Wizard-of-Oz				
Image Caption	A close up of a beige colored couch in floral patterns, with a blue pill bottle in the corner. The pill bottle has blurry small texts.	A close up of a beige colored couch in floral patterns, with a moving, blurry blue object in the corner.	A close up of a beige colored couch in floral patterns, with a black rectangle in the corner.	A close-up of a beige-colored couch in floral patterns, with a blue bottle partially painted in beige in the corner.

Table 5: Example obfuscation applications on the photo (b) (as in Table 3) with descriptions for the obfuscated area (displayed next to its bounding box) and high-level image captions provided by both Wizard-of-Oz and off-the-shelf implementations.

be avoided" (P8). They felt that as long as the caption models were correct and let them know the existence of potentially private objects, they could physically explore and retake the photos to avoid privacy risks (N = 5). For photos participants took themselves, they also generally had a better sense of the objects in the scene and felt more confident in correctly judging the accuracy of the descriptions (N = 6). Other participants, however, were more concerned: *"You can't (always) go back in time and remove it in person. You know, the picture is the picture and you got to be able to remove it after the fact"* (P10).

To mitigate the risks of inaccuracy with automated image editing tools, participants described several possible approaches:

- **Obfuscation Freedom:** Participants commonly desired the freedom to select any object in the photo to be obfuscated (N = 11), given that they did not expect AI to be able to fully accurately detect objects considered private to them across contexts: *"I like having more control over what I'm able to potentially hide"* (P6). Participants wished to further select a specific part of an object—e.g., *"texts"* (P5, P8, P12), *"just the face"* (P11) or a specific photo area with multiple objects: *"maybe divide into 4 or 6 squares...and you could choose to get rid of one"* (P11). P5 and P7 also desired to make the selection directly in the image explorer, while P4 wanted to gradually apply obfuscations layer by layer: *"you*

hid the pill bottle, but it still showed up as a blue thing. What if you could go in and edit it again and hide the blue thing?" (Table 5). While not wanting their obfuscation options to be limited by AI, participants did note that AI could be helpful in reminding them of potentially private objects, given that they themselves could also miss some objects—*"in case I missed it, say hey here're possibly private items... do you want to double check?"* (P9).

- **Multiple Information Sources:** Three participants considered checking the obfuscated photo descriptions with another AI as a way to gain more information and confidence before sharing photos. They thus suggested the tool to incorporate other AI algorithms' assessments for users' easy access: *"building a really quick way to send it to another AI application like, you know, ChatGPT or Be My Eyes"* (P6). Similarly, participants also desired ways to incorporate sighted assistance more effortlessly through the application. In particular, many would like to assess the accuracy of this tool with sighted input prior to using it in real life: *"if I'm editing 10 photos, then if I confirmed with another person and noticed it's making no mistakes...it's like a relationship like you built up the trust"* (P9).
- **Communication about AI Accuracy:** A number of participants also suggested including more information about how well the AI performs to set users' expectations, such as through quantitative measures (e.g., confidence score) (N = 2) and instructions that explain what the tools tend to pick up to encourage critical thinking from the users themselves (as suggested by P6).
- **Improvement of Model Accuracy with Visual Effects:** Besides design improvements, the AI models themselves should consider improving performance in captioning not only clear images but also visual effects, such as blurriness, distortion, and shapes in solid colors (N = 6).

4.3.2 Ineffective Obfuscation Communication. As mentioned in Section 4.1, envisioning obfuscation results can be difficult for blind participants. Descriptions for obfuscation manipulation effects are thus critical. Our prototype communicated obfuscation effects through a high-level summary of the resulting photo and a touch-based exploration of objects identified in the scene. While most participants (N = 8) reacted positively to this approach, we learned that it did not fully support their interpretation needs.

First, participants experienced challenges in locating and evaluating obfuscated areas (N = 9). As suggested in Section 4.2.1, participants were often unsure which description was meant for the obfuscated area and needed to explore through many unaffected objects to arrive at where they were trying to review.

Participants (N = 6) also found that the object descriptions did not provide the most effective information for them to *"determine whether the effect there was applied successfully"* (P6), including to what extent the obfuscation had hidden the private information: *"I wouldn't feel very comfortable because I'm not sure of how blurred out this image is. I'd need some kind of reassurance"* (P12). Participants felt that the mere absence of a description for the private object did not provide enough reassurance: *"it says it's blurry. I'm assuming you can't tell what the pill bottle says, but I don't really like that"* (P3).

Further, participants felt they lacked information about the original photo's composition needed to make informed obfuscation decision, such as objects' relative positions (N = 10)—*"...is the woman sitting in a fifth chair like away from the table, or is she actually at the table?"* (P7), and what was in the foreground of the photo versus background (N = 5). In turn, they were unsure of the objects that would be affected by hiding different obfuscation choices.

Accordingly, we present design insights related to challenges with conveying the output of an obfuscation action:

- **Information critical to interpreting obfuscations:** Our study revealed a set of information key to blind participants' interpretation of obfuscation results, including (1) the visibility of obfuscated private information (N = 7)—e.g., *"if there's any remnant of whatever thing [then] I would need to use a blackout option to hide it better"* (P4); (2) concrete description about the appearance of obfuscation effects, such as just *"blurry"* (P8, P10) or an explicit indication of whether sighted people could identify the information (P4); (3) captions and text detection for objects unaffected by the obfuscated application, as provided in our prototype—all participants found these descriptions helpful in assessing what content had been preserved: *"Very useful cause like I can tell what's in the image and what items I'm hiding" (P7).*
- **Focused description on changes:** Participants desired the descriptions to highlight visual changes (N = 7). For example, one participant suggested: *"Whatever you edited first ends up becoming the first couple of sentences that you hear about an image, because like it makes sense that if you were to look at an edited image, that would still be the first thing that you notice anyways"* (P6). Some further wanted the ability to switch among different obfuscation results: *"this way I could switch out from one view to the next right away to see what I might like and not like"* (P12).
- **Non-visual obfuscation preview:** To better envision obfuscation effects and make informed choices, participants would also like a brief description for what an obfuscation choice entails, prior to actually applying it (N =7). The description should include *"the area that would or wouldn't be affected"* (P12), especially for background obfuscation and bounding boxes: *"give me a better sense of like how big these boxes are"* (P6).

4.3.3 Mixed Feedback on Cognitive Load. Some participants experienced high cognitive load in exploring and obfuscating the photos. For example, P1, P5, and P12 found themselves losing track of previously reviewed obfuscation results after trying out numerous options and navigating the detailed and sometimes inaccurate descriptions for each: *"I chose so many options I couldn't tell anymore what I wanted in what...that part was a little overwhelming"* (P5). Many (N = 6) were concerned about the abundance of captions, especially the repetitive ones caused by algorithm inaccuracy.

At the same time, many participants also felt that the information (N = 6) and options (N = 3) provided in our system were just right: *"It was just enough to get an idea of what the picture was, but not enough to be overwhelmed"* (P5). In fact, despite the concern around

information overload, the majority of participants mentioned at least one additional type of information or functionality they would like the system to provide on top of the existing ones (N = 8), including but not limited to *"colors"* (P12, P9, P3) of objects, *"room decorations"* (P3) and *"people"* (P3, P7, P10), as well as the option to crop a photo (N = 4) and editing photos beyond obfuscation (e.g., photo touch-up, filter) (N = 5). Balancing participants' desire for exploration and concern for cognitive load is therefore a challenge.

Participants mentioned the following relevant design improvements:

- **Minimal design:** Overall, participants valued simplicity (N = 7): *"Our favorite apps are like the ones that got one option. You turn it on and it works"* (P10). They (N = 6) found photos with a smaller number of object captions much easier to interpret and in turn wished to combine repetitive captions—*"a stuffed animal with blue bandana, we know it's (also identified as) a teddy bear...just merge the two descriptions"* (P12)—or describe objects that belong to the same categories in one caption, while clearly indicating the total number and locations of these objects. A number of participants also particularly desired a *"minimal use of sound"* (P4) in reducing their cognitive load (P10, P12, P4): *"SeeingAI makes noise like music when you're sliding your finger around...I find it a little bit annoying to be honest"* (P10).
- **Configurable design:** While the tool design should be overall minimal, it should also accommodate the varied preferences for how much and what information and functions should be provided. Many participants (N = 6) appreciated being able to choose between quickly checking the high-level summary and diving deeper into the photo exploration. Following this approach, they further suggested options to get additional information or functionality as they desire, such as through a *"button where you could get all the details of the image"* (P5), a non-visual *"zoom in"* (P7, P12) function to get more information of a focused area, a *"setting page"* (P9) to personalize information included in photo descriptions (e.g., colors, text identification, people characteristics, position), as well as *"two modes...a photographer mode where you can go in and do fancy touch-ups...then there is a simple (obfuscation) mode"* (P4).

Overall impression: Despite frictions experienced with the prototype, all participants were excited about the overall idea of a screen reader-accessible, AI-powered obfuscation tool—many were eager to use it in everyday life: *"I wish this was actually real that I could take pictures and edit like this"* (P5). Even with existing frictions, some participants felt they would still make use of this prototype in certain ways, such as for more casual scenarios (N = 5) or *"to take a quick picture to send to someone"* (P6). Their urgent desire stems from desires for independence and control over visual content: *"just being able to have more of an awareness of the things that are in image and being able to be more of an active participant in that process...I just really like being able to do stuff like this independently and have an accessible tool that allows you to do it efficiently...since we never really had that opportunity"* (P6). Some commented on the necessity of this tool when sighted help is not available: *"There's a lot of people who don't have someone who can help them edit their pictures"* (P3). Even when needing to check obfuscated results with a sighted person, participants appreciated being able to control the photo themselves first: *"I may elicit a close friend who cited to make sure that the picture didn't have any elements left that shouldn't be shared, but I'll still use it to do most of the editing independently myself before"* (P11). Participants all believed that with design improvements to reduce these frictions, this tool would bring significant positive impact to their life: *"I think this is a program that has very high potential if...you take the suggestions and comments that the participants give you"* (P11).

5 DISCUSSION

Our study explored an AI-assisted obfuscation tool design to support blind individuals in independently controlling and editing private visual content in their photos. Our findings revealed that blind participants were able to use our prototype to interpret and manipulate visual obfuscation details based on considerations related to utility-privacy trade-offs [10, 29, 35], albeit sometimes encountering frictions related to accuracy, non-visual communication, and cognitive load. Participants proposed directions for future design ideas to address these frictions and were hopeful that, with such improvements, AI-assisted obfuscation tools would support their agency in private visual content management. These findings extend prior visual privacy management support for blind individuals (e.g., [8, 87]) with new insights for AI-based tool design to allow more user agency in non-visually manipulating visual privacy obfuscation. Here, we discuss the implications of these findings.

5.1 The Role of AI in Accessible Visual Privacy Obfuscation

Participants generally appreciated the level of user control provided in the prototype. Besides controls recommended by prior interview studies (e.g., dismiss/consent obfuscation) [8, 71], our study shows that options for configuring obfuscation styles and control over the obfuscation area can help blind users manipulate images to meet needs across private content types, recipients, and visual presentation needs. Further controls could even be useful, such as being able to obfuscate any object in the image, adjust obfuscation characteristics (e.g., degree of blurriness), and crop the image. These additional user-initiated manipulation features could help mitigate the risk that users could feel restricted to only AI-based decisions, especially the inaccurate ones. However, more freedom also increases effort and cognitive load. Balancing user agency and effort is thus a non-trivial design goal that requires considering contextual and personal factors.

In turn, we suggest shifting the role of AI in the obfuscation process based on users' needs, adopting Chung et al. [19]'s framework for creative support tool design. Existing obfuscation approaches mainly focus on an *implementation-aiding role* where the AI makes most execution decisions for users (what and how to obfuscate). This approach could benefit individuals who have less desire or capacity to configure obfuscation details in a given situation. However, at the user's command, the system should be able to perform an *evaluation-aiding role* that provides information key to the obfuscation decision (e.g., private object visibility prior and after obfuscation) but not overpowering what the blind user intends to do with

the information. Another important role that has rarely been considered in blind individuals' obfuscation support is *ideation-aiding*. Participants in our study commonly had difficulty envisioning obfuscation effects and would like the system to provide non-visual previews to ease their decisions. Future AI-assisted obfuscation tools should consider including further guidance in these previews, such as ranking options by resulting photo's utility and remnants of private content.

Beyond choosing from these roles, more refined customization could further help meet users' personal needs, such as (a) specifying a user-defined obfuscation style for a specific private object category (e.g., blur for human face, black out for text) but leaving all other decisions to AI, or (b) allowing only some automated obfuscation (e.g., automatic obfuscation of a particularly concerning category) and keeping the AI on an evaluation-aiding role most of the time.

5.2 The Role of Sighted Help in Accessible Visual Privacy Obfuscation

Our findings also revealed a common desire for sighted help to check obfuscation results, especially when participants felt unsure about system accuracy. Blind individuals' collaborative visual privacy management with sighted friends and family is not new [7, 87]. While effective and potentially constructive for interpersonal relationships [87], this approach leaves blind individual's privacy management to the availability and reliability of sighted people and entails potential undesired social cost as well as misalignment in obfuscation goals. Participants in our study commented on the value of being able to make an initial obfuscation attempt before involving sighted assistance, for reasons such as independence, efficiency, privacy concerns, and autonomy.

Based on these insights, future research can reconsider the role of sighted help in visual privacy management. For example, participants emphasized the need to gradually build trust in an AI-assisted obfuscation system—for initial usage, they envisioned needing more sighted help to assess and familiarize themselves with such a tool. Future tool design could consider mechanisms to support such collaborative assessment of the tool, such as providing a record of sighted feedback of a model's performance across different types of photos or a set of example photos for pairs to test and discuss. Beyond this initial learning phase, blind individuals and sighted assistants may find it useful at times to work together on the obfuscation manipulation, for which future tool design should reference mixed-ability collaboration design insights (e.g., [20, 21, 37, 47, 56, 60]).

5.3 Photo Sense-making to Support Obfuscation Manipulation

Our participants made obfuscation decisions based on how they envisioned viewers may interpret the resulting photo (Section 4.2.3), reflecting prior work with sighted users [10, 29, 35]. However, these decisions rely heavily on sense-making of the obfuscated photos, posing more challenges for blind individuals compared to sighted counterparts.

Extending work on image sense-making support for blind people (e.g., [38, 49, 70]), we note some needs similar to general photo exploration (e.g., inclusion of spatial information in caption, hierarchical access to photo content, variation in visual information wants, preference for objects presented as a list) [38, 70], but also other needs unique to private photo obfuscation. In particular, participants needed descriptions beyond object labels—they desired concrete information about the visibility status of the private object and visual appearance of the obfuscated area—which existing tools for visual interpretation fail to support. Computer vision models are known to work less accurately with blurry or dark photos taken by blind individuals [26], and in our study, obfuscated areas often resulted in new false positive object detections. One solution could be to develop models or pipelines that are able to identify obviously obfuscated areas (e.g., blurred or blacked out) rather than attempting to classify those pixels as a non-obfuscated object. These areas could also be described, for example, as "blurry" or otherwise manipulated.

Future tool design should also consider better guiding blind users to understand the results of an obfuscation, such as by summarizing the differences between the obfuscated and original images—suggested by our participants and in prior work [56], using multimodalities to facilitate visual change perceptions [68], and letting users switch between a number of versions quickly to make the contrasts between different options more salient. Regarding the varied information-wants people may have for an obfuscated photo area, involving a visual question answering mechanism could be particularly helpful [71].

5.4 Accessible Image Editing Beyond Obfuscation

Our participants showed strong interest in using features of our prototype for general image editing, echoing interest from the blind community in visual content consumption and creation (e.g., [40, 63, 86]). To date, research on non-visual photo editing support has still been sparse [13, 54, 81]. Some of our design recommendations could apply directly to this general image editing space—such as providing non-visual previews for different visual effects and caution around cognitive demand, though other needs would likely differ. For example, many participants were interested in aesthetic photo touch-up, for which feedback around an edited photo's artistic characteristics, such as styling, mood, angle and lighting, as well as the appearance of the focal figure (e.g., person, animal, scenery) would likely matter more compared to what is needed for visual privacy obfuscation. Towards this extension, future research could consider existing general visual art description guidelines [40] and explore how such guidelines may or may not apply from an editor's perspective. This knowledge would be critical to future development of AI as well as training of sighted help for assisting photo review and editing.

5.5 Limitations

Because of the early stage of this research, our paper focused on a qualitative, exploratory study, using a preliminary prototype design that relied on pre-processing photos. This approach inherently limited what tasks our participants could do, including what photos they obfuscated and the types of objects surfaced to them. Although this approach allowed us to gain an understanding of initial reactions to such AI-based support, these insights may or may not generalize to use in the field. For example, all photos in

our study contained only one private object, due to characteristics of available private visual dataset [66], which limited us from exploring design considerations relevant to situations where multiple items need to be obfuscated (e.g., ranking and categorizing detected private objects). Future studies should consider building a higher-fidelity prototype to further examine the effectiveness of AI-assisted visual obfuscation tools and related design considerations including and beyond the ones proposed in our work. To build such a prototype, technical innovations in the underlying computer vision models are necessary towards better processing of private photos taken by blind people, innovating multimodal models to segment and edit user-specified visual content, as well as computational optimization that allows on-device computation for users' privacy preservation. Future work could also consider exploring ways to refine the Caption-Anything & ChatGPT private object detection approach, such as by including the OCR result in object captions for ChatGPT to process or incorporating alternative large language models to enhance the detection of captions relevant to different privacy categories. Further, we did not obtain an in-depth understanding of participants' past photo editing and obfuscation experience (e.g., editing tool usage) which would likely affect their reactions to new image editing tools. We encourage future studies to further situate accessible visual privacy obfuscation tool design in blind individuals' first-hand experiences. Last, all authors of this work are sighted and could have potentially brought bias to the design and research practices. We practice reflexivity [64] and have sought to center design ideas from blind individuals' perspectives.

6 CONCLUSION

In this work, we explored how blind individuals react to and make use of a preliminary prototype design for obfuscating private visual content in their photos. Through 12 user studies, we uncovered blind participants' mental models and usage patterns with private photo content manipulations, factors that influenced their obfuscation decisions, frictions they experienced with the prototype design, and their design feedback on this line of tools. Overall, participants were excited for potential opportunities to gain more control on visual privacy through this tool, though they emphasized on the importance of reducing frictions related to inaccuracy, poor obfuscation descriptions, and cognitive load. Their specific design ideas inform future accessibility design and computer vision research to reconsider the roles of AI and human assistance as well as alternative visual description practices and model development in supporting accessible photo editing, for visual privacy preservation and beyond.

ACKNOWLEDGMENTS

We thank Jarek Reynolds for his assistance with our prototype implementation and Steven M. Goodman for his feedback on the user study. We also thank our participants for their thoughtful insights. This work was partially supported by the National Science Foundation SaTC grants (#2125925, #2148080, #2126314), the University of Washington's CREATE Center, and Apple Inc. Any views, opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and should not be interpreted as reflecting the views, policies or position, either expressed or implied, of Apple Inc.

REFERENCES

[1] Aira Tech Corp. 2023. *Aira.* Aira Tech Corp. https://aira.io/ Accessed: 2023-9-02.
[2] 2023. Be My Eyes. https://www.bemyeyes.com/. Accessed: 2023-9-02.
[3] Dustin Adams, Lourdes Morales, and Sri Kurniawan. 2013. A Qualitative Study to Support a Blind Photography Mobile Application. In *Proceedings of the 6th International Conference on PErvasive Technologies Related to Assistive Environments* (Rhodes, Greece) *(PETRA '13)*. Association for Computing Machinery, New York, NY, USA, Article 25, 8 pages. https://doi.org/10.1145/2504335.2504360
[4] Tousif Ahmed, Roberto Hoyle, Kay Connelly, David Crandall, and Apu Kapadia. 2015. Privacy Concerns and Behaviors of People with Visual Impairments. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 3523–3532. https://doi.org/10.1145/2702123.2702334
[5] Tousif Ahmed, Patrick Shaffer, Kay Connelly, David Crandall, and Apu Kapadia. 2016. Addressing Physical Safety, Security, and Privacy for People with Visual Impairments. In *Proceedings of the Twelfth USENIX Conference on Usable Privacy and Security* (Denver, CO, USA) *(SOUPS '16)*. USENIX Association, USA, 341–354.
[6] Taslima Akter, Tousif Ahmed, Apu Kapadia, and Manohar Swaminathan. 2022. Shared Privacy Concerns of the Visually Impaired and Sighted Bystanders with Camera-Based Assistive Technologies. *ACM Trans. Access. Comput.* 15, 2, Article 11 (may 2022), 33 pages. https://doi.org/10.1145/3506857
[7] Taslima Akter, Bryan Dosono, Tousif Ahmed, Apu Kapadia, and Bryan Semaan. 2020. "I am uncomfortable sharing what I can't see": Privacy Concerns of the Visually Impaired with Camera Based Assistive Applications. In *29th USENIX Security Symposium (USENIX Security 20)*. USENIX Association, 1929–1948. https://www.usenix.org/conference/usenixsecurity20/presentation/akter
[8] Rahaf Alharbi, Robin N. Brewer, and Sarita Schoenebeck. 2022. Understanding Emerging Obfuscation Technologies in Visual Description Services for Blind and Low Vision People. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2, Article 469 (nov 2022), 33 pages. https://doi.org/10.1145/3555570
[9] Rawan Alharbi, Tammy Stump, Nilofar Vafaie, Angela Pfammatter, Bonnie Spring, and Nabil Alshurafa. 2018. I can't be myself: effects of wearable cameras on the capture of authentic behavior in the wild. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 2, 3 (2018), 1–40.
[10] Rawan Alharbi, Mariam Tolba, Lucia C. Petito, Josiah Hester, and Nabil Alshurafa. 2019. To Mask or Not to Mask? Balancing Privacy with Visual Confirmation Utility in Activity-Oriented Wearable Cameras. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 3, Article 72 (sep 2019), 29 pages. https://doi.org/10.1145/3351230
[11] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N. Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz. 2019. Guidelines for Human-AI Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300233
[12] Mary Jean Amon, Aaron Necaise, Nika Kartvelishvili, Aneka Williams, Yan Solihin, and Apu Kapadia. 2023. Modeling User Characteristics Associated with Interdependent Privacy Perceptions on Social Media. *ACM Trans. Comput.-Hum. Interact.* 30, 3, Article 40 (jun 2023), 32 pages. https://doi.org/10.1145/3577014
[13] 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 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3173650
[14] Danielle Bragg, Oscar Koller, Naomi Caselli, and William Thies. 2020. Exploring Collection of Sign Language Datasets: Privacy, Participation, and Model Performance. In *Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility* (Virtual Event, Greece) *(ASSETS '20)*. Association for Computing Machinery, New York, NY, USA, Article 33, 14 pages. https://doi.org/10.1145/3373625.3417024
[15] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.
[16] Finn Brunton and Helen Nissenbaum. 2015. *Obfuscation: A User's Guide for Privacy and Protest.* The MIT Press.
[17] Diagram Center. [n. d.]. Specific Guidelines: Art, Photos & Cartoons. http://diagramcenter.org/specific-guidelines-final-draft.html. Accessed: 2021-12-02.
[18] Markus Christen, Bert Gordijn, and Michele Loi. 2020. *The ethics of cybersecurity.* Springer Nature.
[19] John Joon Young Chung, Shiqing He, and Eytan Adar. 2021. The Intersection of Users, Roles, Interactions, and Technologies in Creativity Support Tools. In *Proceedings of the 2021 ACM Designing Interactive Systems Conference* (Virtual Event, USA) *(DIS '21)*. Association for Computing Machinery, New York, NY, USA, 1817–1833. https://doi.org/10.1145/3461778.3462050

[20] 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. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 191 (nov 2019), 26 pages. https://doi.org/10.1145/3359293

[21] Maitraye Das, Anne Marie Piper, and Darren Gergle. 2022. Design and Evaluation of Accessible Collaborative Writing Techniques for People with Vision Impairments. *ACM Trans. Comput.-Hum. Interact.* 29, 2, Article 9 (jan 2022), 42 pages. https://doi.org/10.1145/3480169

[22] Pardis Emami-Naeini, Yuvraj Agarwal, Lorrie Faith Cranor, and Hanan Hibshi. 2020. Ask the Experts: What Should Be on an IoT Privacy and Security Label?. In *2020 IEEE Symposium on Security and Privacy (SP)*. 447–464. https://doi.org/10.1109/SP40000.2020.00043

[23] Liang Fang, Lihua Yin, Qiaoduo Zhang, Fenghua Li, and Binxing Fang. 2017. Who Is Visible: Resolving Access Policy Conflicts in Online Social Networks. In *GLOBECOM 2017 - 2017 IEEE Global Communications Conference* (Singapore). IEEE Press, 1–6. https://doi.org/10.1109/GLOCOM.2017.8254015

[24] Mirko Franco. 2023. Toward Safer Social Media Platforms. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 505, 4 pages. https://doi.org/10.1145/3544549.3577059

[25] Ricardo E. Gonzalez Penuela, Paul Vermette, Zihan Yan, Cheng Zhang, Keith Vertanen, and Shiri Azenkot. 2022. Understanding How People with Visual Impairments Take Selfies: Experiences and Challenges. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility* (Athens, Greece) *(ASSETS '22)*. Association for Computing Machinery, New York, NY, USA, Article 63, 4 pages. https://doi.org/10.1145/3517428.3550372

[26] Danna Gurari, Qing Li, Chi Lin, Yinan Zhao, Anhong Guo, Abigale Stangl, and Jeffrey P. Bigham. 2019. VizWiz-Priv: A Dataset for Recognizing the Presence and Purpose of Private Visual Information in Images Taken by Blind People. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

[27] 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 2021 AAAI/ACM Conference on AI, Ethics, and Society* (Virtual Event, USA) *(AIES '21)*. Association for Computing Machinery, New York, NY, USA, 543–554. https://doi.org/10.1145/3461702.3462620

[28] Susumu Harada, Daisuke Sato, Dustin W. Adams, Sri Kurniawan, Hironobu Takagi, and Chieko Asakawa. 2013. Accessible Photo Album: Enhancing the Photo Sharing Experience for People with Visual Impairment. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI '13)*. Association for Computing Machinery, New York, NY, USA, 2127–2136. https://doi.org/10.1145/2470654.2481292

[29] Rakibul Hasan, Eman Hassan, Yifang Li, Kelly Caine, David J. Crandall, Roberto Hoyle, and Apu Kapadia. 2018. Viewer Experience of Obscuring Scene Elements in Photos to Enhance Privacy. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3173621

[30] Rakibul Hasan, Yifang Li, Eman Hassan, Kelly Caine, David J Crandall, Roberto Hoyle, and Apu Kapadia. 2019. Can privacy be satisfying? on improving viewer satisfaction for privacy-enhanced photos using aesthetic transforms. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–13.

[31] Jordan Hayes, Smirity Kaushik, Charlotte Emily Price, and Yang Wang. 2019. Cooperative Privacy and Security: Learning from People with Visual Impairments and Their Allies. In *Fifteenth Symposium on Usable Privacy and Security (SOUPS 2019)*. USENIX Association, Santa Clara, CA, 1–20. https://www.usenix.org/conference/soups2019/presentation/hayes

[32] Panagiotis Ilia, Iasonas Polakis, Elias Athanasopoulos, Federico Maggi, and Sotiris Ioannidis. 2015. Face/Off: Preventing Privacy Leakage From Photos in Social Networks. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security* (Denver, Colorado, USA) *(CCS '15)*. Association for Computing Machinery, New York, NY, USA, 781–792. https://doi.org/10.1145/2810103.2813603

[33] Chandrika Jayant, Hanjie Ji, Samuel White, and Jeffrey P. Bigham. 2011. Supporting Blind Photography *(ASSETS '11)*. Association for Computing Machinery, New York, NY, USA, 203–210. https://doi.org/10.1145/2049536.2049573

[34] 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 *Proceedings of the 2022 ACM Designing Interactive Systems Conference* (Virtual Event, Australia) *(DIS '22)*. Association for Computing Machinery, New York, NY, USA, 1893–1906. https://doi.org/10.1145/3532106.3533522

[35] Sanjay Kairam, Joseph 'Jofish' Kaye, John Alexis Guerra-Gomez, and David A. Shamma. 2016. Snap Decisions? How Users, Content, and Aesthetics Interact to Shape Photo Sharing Behaviors. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 113–124. https://doi.org/10.1145/2858036.2858451

[36] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick. 2023. Segment Anything. arXiv:2304.02643 [cs.CV]

[37] Hae-Na Lee, Yash Prakash, Mohan Sunkara, I.V. Ramakrishnan, and Vikas Ashok. 2022. Enabling Convenient Online Collaborative Writing for Low Vision Screen Magnifier Users. In *Proceedings of the 33rd ACM Conference on Hypertext and Social Media* (Barcelona, Spain) *(HT '22)*. Association for Computing Machinery, New York, NY, USA, 143–153. https://doi.org/10.1145/3511095.3531274

[38] Jaewook Lee, Jaylin Herskovitz, Yi-Hao Peng, and Anhong Guo. 2022. ImageExplorer: Multi-Layered Touch Exploration to Encourage Skepticism Towards Imperfect AI-Generated Image Captions. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 462, 15 pages. https://doi.org/10.1145/3491102.3501966

[39] Fenghua Li, Zhe Sun, Ben Niu, Jin Cao, and Hui Li. 2019. An Extended Control Framework for Privacy-Preserving Photo Sharing across Different Social Networks. In *2019 International Conference on Computing, Networking and Communications (ICNC)*. 390–394. https://doi.org/10.1109/ICCNC.2019.8685488

[40] Franklin Mingzhe Li, Lotus Zhang, Maryam Bandukda, Abigale Stangl, Kristen Shinohara, Leah Findlater, and Patrick Carrington. 2023. Understanding Visual Arts Experiences of Blind People. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 60, 21 pages. https://doi.org/10.1145/3544548.3580941

[41] Jingyi Li, Son Kim, Joshua A. Miele, Maneesh Agrawala, and Sean Follmer. 2019. Editing Spatial Layouts through Tactile Templates for People with Visual Impairments. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3290605.3300436

[42] Junnan Li, Dongxu Li, Silvio Savarese, and Steven Hoi. 2023. BLIP-2: Bootstrapping Language-Image Pre-training with Frozen Image Encoders and Large Language Models. arXiv:2301.12597 [cs.CV]

[43] Wenjie Li, Rongrong Ni, and Yao Zhao. 2017. JPEG photo privacy-preserving algorithm based on sparse representation and data hiding. In *Image and Graphics: 9th International Conference, ICIG 2017, Shanghai, China, September 13-15, 2017, Revised Selected Papers, Part III 9*. Springer, 575–586.

[44] Yifang Li and Kelly Caine. 2022. Obfuscation Remedies Harms Arising from Content Flagging of Photos. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 35, 25 pages. https://doi.org/10.1145/3491102.3517520

[45] Yifang Li, Nishant Vishwamitra, Hongxin Hu, and Kelly Caine. 2020. Towards A Taxonomy of Content Sensitivity and Sharing Preferences for Photos. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376498

[46] Yifang Li, Nishant Vishwamitra, Bart P. Knijnenburg, Hongxin Hu, and Kelly Caine. 2017. Effectiveness and Users' Experience of Obfuscation as a Privacy-Enhancing Technology for Sharing Photos. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW, Article 67 (dec 2017), 24 pages. https://doi.org/10.1145/3134702

[47] Guilherme H. M. Marques, Daniel C. Einloft, Augusto C. P. Bergamin, Joice A. Marek, Renan G. Maidana, Marcia B. Campos, Isabel H. Manssour, and Alexandre M. Amory. 2017. Donnie robot: Towards an accessible and educational robot for visually impaired people. In *2017 Latin American Robotics Symposium (LARS) and 2017 Brazilian Symposium on Robotics (SBR)*. 1–6. https://doi.org/10.1109/SBR-LARS-R.2017.8215273

[48] Microsoft. 2023. *Microsoft Computer Vision API Documentation*. Microsoft Corporation. https://learn.microsoft.com/en-us/azure/ai-services/computer-vision/concept-ocr Accessed: 2023-7-12.

[49] Meredith Ringel Morris, Jazette Johnson, Cynthia L. Bennett, and Edward Cutrell. 2018. Rich Representations of Visual Content for Screen Reader Users. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3173574.3173633

[50] Helen Nissenbaum. 2004. Privacy as contextual integrity. *Wash. L. Rev.* 79 (2004), 119.

[51] Angelo Nodari, Marco Vanetti, and Ignazio Gallo. 2012. Digital privacy: Replacing pedestrians from google street view images. In *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*. IEEE, 2889–2893.

[52] OpenAI. 2022. ChatGPT: A Language Model by OpenAI. https://openai.com/blog/chatgpt

[53] José Ramón Padilla-López, Alexandros Andre Chaaraoui, Feng Gu, and Francisco Flórez-Revuelta. 2015. Visual privacy by context: proposal and evaluation of a level-based visualisation scheme. *Sensors* 15, 6 (2015), 12959–12982.

[54] Soobin Park. 2020. Supporting Selfie Editing Experiences for People with Visual Impairments. In *Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility* (Virtual Event, Greece) *(ASSETS '20)*. Association for Computing Machinery, New York, NY, USA, Article 106, 3 pages. https:

//doi.org/10.1145/3373625.3417082
[55] Deepak Pathak, Philipp Krahenbuhl, Jeff Donahue, Trevor Darrell, and Alexei A. Efros. 2016. Context Encoders: Feature Learning by Inpainting. arXiv:1604.07379 [cs.CV]
[56] Yi-Hao Peng, Jason Wu, Jeffrey Bigham, and Amy Pavel. 2022. Diffscriber: Describing Visual Design Changes to Support Mixed-Ability Collaborative Presentation Authoring. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (Bend, OR, USA) *(UIST '22)*. Association for Computing Machinery, New York, NY, USA, Article 35, 13 pages. https://doi.org/10.1145/3526113.3545637
[57] Helen Petrie, Chandra Harrison, and Sundeep Dev. 2005. Describing images on the web: a survey of current practice and prospects for the future. *Proceedings of Human Computer Interaction International (HCII)* 71, 2 (2005).
[58] Pexels. 2023. *Pexels - Free Stock Photos and Videos*. https://www.pexels.com/ Accessed: 2023-9-02.
[59] Venkatesh Potluri, Tadashi Grindeland, Jon E Froehlich, and Jennifer Mankoff. 2019. Ai-assisted ui design for blind and low-vision creators. In *the ASSETS'19 Workshop: AI Fairness for People with Disabilities*.
[60] Venkatesh Potluri, Maulishree Pandey, Andrew Begel, Michael Barnett, and Scott Reitherman. 2022. CodeWalk: Facilitating Shared Awareness in Mixed-Ability Collaborative Software Development. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility* (Athens, Greece) *(ASSETS '22)*. Association for Computing Machinery, New York, NY, USA, Article 20, 16 pages. https://doi.org/10.1145/3517428.3544812
[61] Chi-Hyoung Rhee and Chang Ha Lee. 2013. Cartoon-like Avatar Generation Using Facial Component Matching. https://api.semanticscholar.org/CorpusID:10274912
[62] Sanster. 2023. lama-cleaner: A Tool for Cleaning and Optimizing LLMs. https://github.com/Sanster/lama-cleaner. Accessed: 2023-9-02.
[63] Anastasia Schaadhardt, Alexis Hiniker, and Jacob O. Wobbrock. 2021. Understanding Blind Screen-Reader Users' Experiences of Digital Artboards. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 270, 19 pages. https://doi.org/10.1145/3411764.3445242
[64] Stephen Secules, Cassandra McCall, Joel Alejandro Mejia, Chanel Beebe, Adam S. Masters, Matilde L. Sánchez-Peña, and Martina Svyantek. 2021. Positionality practices and dimensions of impact on equity research: A collaborative inquiry and call to the community. *Journal of Engineering Education* 110, 1 (2021), 19–43. https://doi.org/10.1002/jee.20377 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/jee.20377
[65] Sadia Shamma and Md Yusuf Sarwar Uddin. 2014. Towards privacy-aware photo sharing using mobile phones. In *8th International Conference on Electrical and Computer Engineering*. 449–452. https://doi.org/10.1109/ICECE.2014.7026919
[66] Tanusree Sharma, Abigale Stangl, Lotus Zhang, Yu-Yun Tseng, Inan Xu, Leah Findlater, Danna Gurari, and Yang Wang. 2023. Disability-First Design and Creation of A Dataset Showing Private Visual Information Collected With People Who Are Blind. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 51, 15 pages. https://doi.org/10.1145/3544548.3580922
[67] Jiayu Shu, Rui Zheng, and Pan Hui. 2018. Cardea: Context-Aware Visual Privacy Protection for Photo Taking and Sharing. In *Proceedings of the 9th ACM Multimedia Systems Conference* (Amsterdam, Netherlands) *(MMSys '18)*. Association for Computing Machinery, New York, NY, USA, 304–315. https://doi.org/10.1145/3204949.3204973
[68] Taliesin L. Smith and Emily B. Moore. 2020. Storytelling to Sensemaking: A Systematic Framework for Designing Auditory Description Display for Interactives. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376460
[69] Agrima Srivastava and G Geethakumari. 2014. A privacy settings recommender system for Online Social Networks. In *International Conference on Recent Advances and Innovations in Engineering (ICRAIE-2014)*. 1–6. https://doi.org/10.1109/ICRAIE.2014.6909142
[70] Abigale Stangl, Shasta Ihorn, Yue-Ting Siu, Aditya Bodi, Mar Castanon, Lothar Narins, and Ilmi Yoon. 2023. The Potential of a Visual Dialogue Agent In a Tandem Automated Audio Description System for Videos. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–16.
[71] Abigale Stangl, Emma Sadjo, Pardis Emami-Naeini, Yang Wang, Danna Gurari, and Leah Findlater. 2023. "Dump It, Destroy It, Send It to Data Heaven": Blind People's Expectations for Visual Privacy in Visual Assistance Technologies. In *Proceedings of the 20th International Web for All Conference* (Austin, TX, USA) *(W4A '23)*. Association for Computing Machinery, New York, NY, USA, 134–147. https://doi.org/10.1145/3587281.3587296
[72] Abigale Stangl, Kristina Shiroma, Nathan Davis, Bo Xie, Kenneth R. Fleischmann, Leah Findlater, and Danna Gurari. 2022. Privacy Concerns for Visual Assistance Technologies. *ACM Trans. Access. Comput.* 15, 2, Article 15 (may 2022), 43 pages. https://doi.org/10.1145/3517384
[73] Abigale Stangl, Kristina Shiroma, Bo Xie, Kenneth R. Fleischmann, and Danna Gurari. 2020. Visual Content Considered Private by People Who Are Blind. In *Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility* (Virtual Event, Greece) *(ASSETS '20)*. Association for Computing Machinery, New York, NY, USA, Article 31, 12 pages. https://doi.org/10.1145/3373625.3417014
[74] 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 23rd International ACM SIGACCESS Conference on Computers and Accessibility* (Virtual Event, USA) *(ASSETS '21)*. Association for Computing Machinery, New York, NY, USA, Article 16, 15 pages. https://doi.org/10.1145/3441852.3471233
[75] Jose M. Such, Joel Porter, Sören Preibusch, and Adam Joinson. 2017. Photo Privacy Conflicts in Social Media: A Large-Scale Empirical Study. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 3821–3832. https://doi.org/10.1145/3025453.3025668
[76] Weiwei Sun, Jiantao Zhou, Ran Lyu, and Shuyuan Zhu. 2016. Processing-Aware Privacy-Preserving Photo Sharing over Online Social Networks. In *Proceedings of the 24th ACM International Conference on Multimedia* (Amsterdam, The Netherlands) *(MM '16)*. Association for Computing Machinery, New York, NY, USA, 581–585. https://doi.org/10.1145/2964284.2967288
[77] Roman Suvorov, Elizaveta Logacheva, Anton Mashikhin, Anastasia Remizova, Arsenii Ashukha, Aleksei Silvestrov, Naejin Kong, Harshith Goka, Kiwoong Park, and Victor Lempitsky. 2021. Resolution-robust Large Mask Inpainting with Fourier Convolutions. arXiv:2109.07161 [cs.CV]
[78] Madiha Tabassum, Abdulmajeed Alqhatani, Marran Aldossari, and Heather Richter Lipford. 2018. Increasing User Attention with a Comic-Based Policy. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3173574.3173774
[79] Matt Tierney, Ian Spiro, Christoph Bregler, and Lakshminarayanan Subramanian. 2013. Cryptagram: Photo Privacy for Online Social Media. In *Proceedings of the First ACM Conference on Online Social Networks* (Boston, Massachusetts, USA) *(COSN '13)*. Association for Computing Machinery, New York, NY, USA, 75–88. https://doi.org/10.1145/2512938.2512939
[80] Nishant Vishwamitra, Yifang Li, Kevin Wang, Hongxin Hu, Kelly Caine, and Gail-Joon Ahn. 2017. Towards PII-Based Multiparty Access Control for Photo Sharing in Online Social Networks. In *Proceedings of the 22nd ACM on Symposium on Access Control Models and Technologies* (Indianapolis, Indiana, USA) *(SACMAT '17 Abstracts)*. Association for Computing Machinery, New York, NY, USA, 155–166. https://doi.org/10.1145/3078861.3078875
[81] 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 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (San Francisco, California, USA) *(CSCW '16)*. Association for Computing Machinery, New York, NY, USA, 1584–1595. https://doi.org/10.1145/2818048.2820013
[82] Teng Wang, Jinrui Zhang, Junjie Fei, Yixiao Ge, Hao Zheng, Yunlong Tang, Zhe Li, Mingqi Gao, Shanshan Zhao, Ying Shan, and Feng Zheng. 2023. Caption anything: Interactive image description with diverse multimodal controls. *arXiv preprint arXiv:2305.02677* (2023).
[83] Shaomei Wu, Jeffrey Wieland, Omid Farivar, and Julie Schiller. 2017. Automatic alt-text: Computer-generated image descriptions for blind users on a social network service. In *proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*. 1180–1192.
[84] Kaihe Xu, Yuanxiong Guo, Linke Guo, Yuguang Fang, and Xiaolin Li. 2017. My Privacy My Decision: Control of Photo Sharing on Online Social Networks. *IEEE Transactions on Dependable and Secure Computing* 14, 2 (2017), 199–210. https://doi.org/10.1109/TDSC.2015.2443795
[85] Lin Yuan, Pavel Korshunov, and Touradj Ebrahimi. 2015. Secure JPEG scrambling enabling privacy in photo sharing. In *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, Vol. 04. 1–6. https://doi.org/10.1109/FG.2015.7285022
[86] Lotus Zhang, Simon Sun, and Leah Findlater. 2023. Understanding Digital Content Creation Needs of Blind and Low Vision People. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–15.
[87] Zhuohao (Jerry) Zhang, Smirity Kaushik, JooYoung Seo, Haolin Yuan, Sauvik Das, Leah Findlater, Danna Gurari, Abigale Stangl, and Yang Wang. 2023. ImageAlly: A Human-AI Hybrid Approach to Support Blind People in Detecting and Redacting Private Image Content. In *Nineteenth Symposium on Usable Privacy and Security (SOUPS 2023)*. USENIX Association, Anaheim, CA, 417–436. https://www.usenix.org/conference/soups2023/presentation/zhang
[88] 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. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW, Article 121 (dec 2017), 22 pages. https://doi.org/10.1145/3134756

[89] Yuhang Zhao, Shaomei Wu, Lindsay Reynolds, and Shiri Azenkot. 2018. A Face Recognition Application for People with Visual Impairments: Understanding Use Beyond the Lab. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3173574.3173789