VizXpress: Towards Expressive Visual Content by Blind Creators Through AI Support

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

Zhuohao (Jerry) Zhang
University of Washington
Seattle, WA, USA
zhuohao@uw.edu

Gina Clepper
University of Washington
Seattle, WA, USA
gclepper@uw.edu

Franklin Mingzhe Li
Carnegie Mellon University
Pittsburgh, PA, USA
mingzhe2@cs.cmu.edu

Patrick Carrington
Carnegie Mellon University
Pittsburgh, PA, USA
pcarrington@cmu.edu

Jacob O. Wobbrock
University of Washington
Seattle, WA, USA
wobbrock@uw.edu

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

Abstract

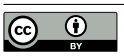
From curating the layout of a resume to selecting filters for social media, creating and configuring visual content allows individuals to express identity, communicate intent, and engage socially, yet blind individuals often face significant barriers to such expressive practices. Prior accessibility research primarily addresses *functional* content configuration, leaving little understanding of blind individuals' *expressive* visual creation needs. To better understand and support these needs, we conducted a two-stage study: first, we interviewed 10 blind participants to understand their motivations, current practices, and barriers in visual expression, and to ideate on potential visual editing support; second, based on interview insights, we developed an interactive prototype (VizXpress) that provides real-time feedback on visual aesthetics using a vision-language model and supports automated and manual visual editing controls. We used VizXpress as a design probe to further explore accessible design opportunities for visual expression. Our findings highlight many blind users' strong interest in creating visually expressive content, nuanced informational requirements for subjective aesthetics (e.g., color, mood, lighting), and ongoing accessibility challenges with visual creative tools. Grounded in these insights, we propose design implications including richer aesthetic feedback, controlled intelligent editing, and accessible manual editing mechanisms.

CCS Concepts

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

Keywords

accessibility, creativity support



ASSETS '25, Denver, CO, USA

ACM ISBN 979-8-4007-0676-9/25/10
https://doi.org/10.1145/3663547.3746345

ACM Reference Format:
Lotus Zhang, Zhuohao (Jerry) Zhang, Gina Clepper, Franklin Mingzhe Li, Patrick Carrington, Jacob O. Wobbrock, and Leah Findlater. 2025. VizXpress: Towards Expressive Visual Content by Blind Creators Through AI Support. In *The 27th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '25), October 26–29, 2025, Denver, CO, USA.* ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3663547.3746345

1 Introduction

Digital visual content sharing often involves intentional aesthetic choices to express oneself and connect with others—whether selecting filters for social media, curating the layout of a resume, or choosing the visual theme of a presentation. These expressive visual edits can be deeply tied to how people manage impressions, communicate identity, and convey intent, and they can also be a meaningful source of personal enjoyment and artistic exploration. Past work has demonstrated the interest that many blind individuals have in these creative, social, and self-expressive practices [13, 125]. However, access to digital visual expressions has been limited for blind users, restricting participation for those who are interested [90, 101, 125].

Accessibility researchers have increasingly explored how to support blind users in configuring visual content [20, 43, 49, 62, 124, 127]. Much of this work has focused on making visual configuration more *functionally* accessible, supporting tasks such as privacy preservation [124] and document formatting [127]. These approaches often prioritize task completion and utility, rather than expressiveness or aesthetics. A small portion of this research that aims to broaden blind users' creative control has focused on technological explorations of new non-visual interaction techniques and their effectiveness in supporting object-level modifications [20, 62]. While this research provides helpful design insights, blind individuals' access to visual creation has been limited to a small set of edit actions, leaving more *expressive* options inaccessible. So far, there remains a lack of understanding and support for blind individuals' desires for more expressive freedom with digital visual media.

To better support blind individuals' creative needs, we conducted a two-stage study aimed at understanding and expanding opportunities for accessible visual expression. We centered our research on the following questions:

- RQ1: What interests do blind individuals have in expressive visual creation, if any?
- RQ2: How do blind individuals currently engage with expressive visual creation?
- RQ3: How should creative tools be designed to support their expressive needs?

We first conducted an interview study with 10 blind participants to explore their needs, perspectives, and experiences with creative visual expression and to collaboratively ideate on design directions for future tool support. The interview revealed interest in specific aspects of digital visual expressions (e.g., aesthetics and styling, visual effects and overlays, illustrations and art), motivations behind these interests (e.g., engaging others, professional goals, artistic pursuits), and design implications for under-supported needs: (1) richer, actionable, and aesthetics-related visual feedback, (2) guidance and automation for accessible visual configuration, and (3) support for creator agency.

We translated our interview findings into an interactive prototype, VizXpress, that provides: (1) real-time feedback on image aesthetics, (2) automated editing based on users' creative intents, and (3) accessible manual editing controls with screen-reader compatibility. We used VizXpress as a design probe in a study with 14 blind participants to further investigate how tool design could support their digital visual expression. The design probe study revealed blind users' varied information needs for editing images across contexts, challenges with perceiving nuanced or subjective visual changes (e.g., color, lighting, vibe), and design insights for screen-reader friendly manual visual edits (e.g., object-based cropping, descriptive visual styling options). From this research, we identified promising directions to allow greater access and freedom in blind individuals' visual creation.

Taken together, this research contributes: (1) an in-depth understanding of blind individuals' experience and interests with expressive visual creation; (2) a set of co-ideated and evaluated design requirements for creative tools to enable greater access and freedom in blind users' digital visual expressions; (3) a prototype to support blind users' expressive image editing, using AI-based aesthetics feedback and accessible edit controls.

2 Related Work

Our work builds on prior research in visual content creation accessibility, screen-reader access to visual content, expression and aesthetics in accessibility, and AI-based visual content creation.

2.1 Accessibility of Visual Content Creation

Previous research has consistently highlighted interest from the blind community in creating visual digital media, including photography [1, 14, 39, 52, 117], video [49, 98, 103], visual layouts [66, 83, 88, 90, 91, 101], and artistic content [15, 22]. Blind individuals engage with visual creation for various personal, professional, and social reasons [125], including sharing media on social platforms, creating personal keepsakes, and authoring materials such as presentations for work or education [2, 11, 39, 83, 88, 98, 101, 102]. Recent work has also noted interest in more expressive visual creation, especially artistic photography and illustration [13, 48, 125].

Blind creators often encounter significant barriers with evaluating visual outcomes [2, 11, 48, 49, 101, 125], understanding visual concepts [90, 125], and navigating inaccessible creative interfaces [49, 101, 125]. Accessibility research has begun to tackle some of these challenges. Extensive work has focused on blind photography, supporting camera framing, composition, and object capturing [9, 43, 61, 68]. Other studies have looked into post-production tasks, such as obfuscating private images [124], facilitating object-level manipulations via text commands and verification loops [20], and facial retouching [84]. In visual layout design (e.g., presentation slides, user interfaces), research has explored AI-based and multimodal feedback to convey visual structure and detect potential issues [47, 65, 67, 88, 91, 101, 127]. For video editing, similar methods were explored for visual issue detection, along with tool innovations for the temporal aspect (e.g., script-based video timeline navigation [49]). Artistic content creation has also received accessible design attention. Non-visual drawing and illustrations, for example, could benefit from audio-tactile feedback, voice-based interactions, grid- or tile-based spatial control, and text-based content creation through generative AI [35, 40, 54, 59, 62].

So far, these accessibility supports have mainly focused on functional outcomes such as task completion, efficiency and accuracy, limited to defined objectives (e.g., object-level edits, layout issues, privacy preservation). We know little about how blind creators approach expressive or aesthetic visual edits—tasks inherently subjective and closely tied to creative intent, personal expression, and aesthetic judgment. Gaining this understanding becomes particularly relevant as growing evidence indicates expressive needs in visual creation, coupled with unique accessibility barriers in evaluating and applying complex visual aesthetics [11, 33, 63, 84, 98, 125].

2.2 Screen-Reader Access to Visual Content

A substantial body of research exists regarding screen-reader accessibility for visual content consumption. Guidelines have emphasized creating effective descriptions tailored to blind users' informational goals and contexts, prioritizing efficiency, clarity, and relevance [26, 29, 36, 71, 110, 122]. For instance, social media images require descriptive information on personal identities, locations, and viewer reactions [75, 117, 126, 129], while artistic visuals demand descriptions attentive to high-level narratives, subjects, and specific artistic details [21, 46, 64]. For visual layouts such as presentation slides and websites, a clear description to visual elements, hierarchical structure, and context is critical [36, 74, 86, 87]. Similarly, video accessibility guidelines highlight flexible and context-sensitive approaches, as well as the temporal and narrative alignment of audio descriptions [10, 53, 116].

To handle the complexity and volume of visual descriptions, research advocates a hierarchical approach, starting from concise summaries and expanding selectively based on user interest and context [74, 110, 128]. Multimodal supplements (haptic and audio feedback) further improve perceptual accuracy and efficiency, especially for spatial understanding [42, 53, 60, 68, 77, 78, 106, 127, 130]. AI-generated visual feedback has been increasingly adopted and

benefits screen-reader accessibility, though it presents challenges with accuracy, lack of transparency, and over-trust due to limited verification options [3, 6, 33, 44, 129]. Past research thus emphasized the importance of clear explanations and confidence indicators to support non-visual sense-making [6, 25, 44, 122].

Recently, these principles have extended into visual feedback practices for creative contexts, recommending multimodal feedback and hierarchical structures for conveying visual changes, spatial relationships, and potential layout issues [47, 62, 88, 101, 127]. Past work, however, predominantly targets visual feedback relevant to functional edits rather than expressive or aesthetic tasks involving subjective visual decisions such as mood or personal style—a common and important component of visual creation [27, 69, 79]. Our work addresses this overlooked area by examining effective visual feedback mechanisms specific to blind creators' visual expression needs.

2.3 Expression and Aesthetics in Accessibility

The right to aesthetic experiences is fundamental to cultural participation and personal well-being [73, 76, 96, 121]. Literature has long acknowledged many blind individuals' interests in aesthetic engagement [8, 37, 41]. Accessibility researchers have explored non-visual aesthetic experiences across visual art, museums, and digital interfaces through multimodal methods such as audio and tactile access and interactive tool design [4, 7, 46, 64, 74, 95, 109, 123].

Artistic creation, intimately linked with aesthetic appreciation [113], also serves as an important mode of self-expression for blind individuals' identity, agency, and artistic vision, challenging the visual-centric assumptions in art and enriching cultural, aesthetic discourse [18, 57, 100, 111]. Accessibility research exploring disabled artists' practices across various media (crafting, audio and music, visual art) reveals extensive access labor in technology and tool repurposing to enable artistic expression [13, 23, 70, 85, 99, 125]. Still, the lack of accessible visual creative tools limits blind creators' expression and artistic experiences [13, 125]. This paper contributes an understanding of blind creators' experiences and needs when expressing aesthetics visually, as well as their perspectives on potential directions for tool design.

2.4 AI-Assisted Visual Content Creation

Recent advancements in AI, particularly generative and large multimodal models (e.g., [56, 82, 93, 114]), have expanded visual content creation and editing capabilities, facilitating tasks ranging from high-quality visual synthesis to style modifications and object manipulations through natural language for images [17, 51, 81, 97, 107, 115], videos [50, 92, 94, 120], and visual layout design [28, 30, 45]. Prompt engineering has emerged as an effective method to influence and guide model outputs through wording, structure, specificity, and iterative refinement strategies [19, 24, 58]. At the same time, the use of generative AI in creative contexts raises important concerns around accuracy, bias, authenticity, and particularly visual accessibility for blind users, who rely on non-visual methods to interpret and assess AI-generated visual outcomes [31, 47, 71, 104, 105, 112, 119].

Accessibility research has begun examining blind individuals' interaction with AI-based visual creative tools. For instance, prior studies explored AI-based verification loops and iterative feedback strategies supporting object-level edits, layout design, scene construction, and privacy-preserving image obfuscations [20, 47, 48, 62, 88, 124], highlighting the importance of balancing automation and user agency [20, 62, 124]. Nevertheless, existing research has largely focused on structured composition, prompt verification, or functional manipulation of objects and layouts. While some tools enable basic stylistic edits or assist in evaluating visual outputs, they stop short of engaging with the creative process itself—how blind creators develop, explore, and express aesthetic ideas such as mood, tone, or personal style. There is limited understanding of how AI tools may support blind users' expressive visual authorship. This work explores how blind creators approach and react to AI tools for expressive visual creation, in contrast to their existing practices, offering insights to guide accessible AI-assisted creative tool design.

3 Interview Study Method: Understanding Visual Expression Needs of Blind Individuals

To understand the visual expression needs of blind individuals, we first conducted an interview study ($N = 10$). We asked about participants' interests and experiences with creating visual content, followed by discussions on potential technological support through both open-ended brainstorming and feedback elicitation on audio mockups that instantiated a range of AI- and peer-based design ideas.

3.1 Participants

We recruited 10 participants through the National Federation of the Blind mailing list and a recruitment list maintained by our research group. All participants were 18 years of age or older, identified as blind (including completely blind, legally blind, or having some light perception), and primarily used a screen-reader to access technology. While many blind individuals are interested in visual creation, others are not [125]; as such, we recruited participants who confirmed interest in visual expression in a screener. Table 1 presents participants' demographics (I1-I10). Participants were offered a $30 Amazon gift card for 60 minutes of their time.

3.2 Study Procedure

Our interview sessions were hosted remotely via the Zoom video-conferencing platform. After a brief introduction and demographic questions, we asked about interests with authoring visual content and configuring aesthetics and probed on motivations for engaging with specific types of visual content that participants mentioned (e.g., photos, videos, documents): *"Why did you want the content to be aesthetic [or a specific style participants mentioned]? What aspects of the content come to your mind when considering visual aesthetics and style?"* We then asked about any prior experience with visual aesthetics configuration; for participants with experience, we asked for a detailed walkthrough to learn about their process and challenges, whereas for those without prior experience, we asked for things that stopped them from doing so and any example scenarios where they felt a need for it.

Participant	Gender	Age	Visual Condition	Onset	Visual Memory?
I1 (D7)	Woman	55	Totally Blind	18 yo	Yes
I2 (D8)	Woman	29	Some Light Perception	13 yo	Yes
I3	Woman	40	Totally Blind	Birth	Yes
I4	Woman	57	Some Light Perception	16 yo	Yes
I5	Woman	24	Legally Blind	4 yo	Yes
I6 (D1)	Woman	59	Totally Blind	Birth	Yes
I7 (D3)	Woman	32	Totally Blind	1 yo	No
I8	Man	20	Some Light Perception	Birth	No
I9	Man	46	Totally Blind	Birth	No
I10	Man	31	Legally Blind	10 yo	Yes
D2	Woman	47	Legally Blind	9 yo	Yes
D4	Man	50	Totally Blind	Birth	No
D5	Woman	60	Some Light Perception	Birth	No
D6	Woman	39	Legally Blind	Birth	Limited
D9	Woman	75	Totally Blind	Birth	No
D10	Man	56	Totally Blind	Birth	No
D11	Woman	62	Legally Blind	Birth	Limited
D12	Man	30	Some Light Perception	Birth	No
D13	Man	41	Legally Blind	Birth	Limited
D14	Man	34	Legally Blind	Birth	Limited

Table 1: Demographic information for the $N = 10$ participants in the interview study described in Section 3 (marked with an I) and the $N = 14$ participants in the design probe study described in Section 6 (marked with a D). $N = 20$ unique individuals took part in total. The four individuals who participated in both studies are identified with two participant IDs, e.g. I1 and D7.

We then invited participants to envision and critique ideas for making visual creative tools more accessible, starting with an independent brainstorming session: *"If you were to envision some ideal ways to author or edit visual content, what would it be like?"* To ground further discussion and help participants envision future tool capabilities, we presented three design ideas through audio demos: (1) a *visual aesthetics AI guide* that provides information and guidance for configuring visual content but does not directly make edits for the user; (2) an *automatic AI visual refiner* that configures visual content based on a user's request; and (3) a *peer support tool* that connects users with sighted peers for support. We derived these ideas from prior literature that identified blind individuals' needs for general visual creation (e.g., [20, 48, 125]) and an exploration of off-the-shelf models' capabilities to provide visual aesthetic evaluation and editing suggestions. The audio demos each included a brief explanation of how the idea works and three in-scenario, imaginary use cases: (a) presentation slide editing, (b) photo taking, and (c) photo editing, all common tasks desired by blind individuals [125]. The demos also clarified that these ideas can be used for other types of visual content, and that the AI could be inaccurate. (The audio demos are included in Supplementary Materials. All AI-generated feedback in the demos was produced using GPT-4o.) As participants listened to audio demos, they were encouraged to critique the conceptual design and share any suggestions for improvement or new ideas. After all three demos, we again asked for perspectives on an ideal tool to support their visual expression needs and invited any open-ended comments.

3.3 Data Analysis

Interviews were audio recorded and transcribed. To extract participants' interests, challenges, and design requirements, we adopted an exploratory thematic analysis approach [16]. The first author went through all transcripts to develop an initial codebook, which the research team reviewed and revised collaboratively to enhance theme clarity and relevance to research goals. The first author then independently coded all transcripts, with the third author reviewing half of them selected by a random number generator. The iterated final themes included: (1) visual expression interest, (2) motivations, (3) aspects of visual expression deemed important, (4) existing practices, (5) challenges, (6) design requirements, and (6) reactions to design ideas.

4 Interview Findings

We organize our findings into visual expression interests, experiences and challenges, and ideas for future tool improvements.

4.1 Interests in Expressive Visual Creation

Participants wanted to author expressive visual content for a range of motivations. Many ($N = 10$) mentioned professional or educational goals, such as I2, a teacher, who wanted to create *"a teaching blog with engaging pictures"* (I2) for students and peers, and I9, a musician, who sought to advertise for concerts through *"[social video] content to point people back to my music and a flyer for people to come see our show."* Other career-related examples include presentation slides for service dog education, videos of a youth program, a website for a non-profit, product design and images for

an online marketplace, resumes, and business cards. At the same time, many ($N = 7$) expressed an innate *interest in aesthetic and artistic matters*, such as adding *"something fun and decorative"* (I4) to visual documents, or art-making, like I6, who drew cartoons with remaining vision: *"I loved Gary Larson's 'Far Side' series, and that's sort of my aesthetics. Now I kind of substituted those things with other forms of art [...] like taking pictures at night of the moon and astronomical [events]"* (I6). I7, who became blind at a young age, also shared artistic interests: *"Even though I don't have any vision, I love colors and details. I really enjoy taking pictures and feel like art, visual media is a way to connect with people."* Last, participants also wanted to enhance visual aesthetics for social media posting ($N = 7$) and content sharing with family and friends ($N = 6$). Overall, participants felt strongly that the visual appearance of their content impacts their engagement with the sighted and low vision community and considered it key to message delivery: *"If you're trying to attract the attention of somebody listening to heavy-metal, that's gonna have a different visual appeal than somebody who's going to the symphony. Your visuals can add to your message"* (I4).

Participants shared specific interests for different visual media. For photo and video, participants wanted accessible ways to (1) apply visual effects and overlays ($N = 8$) and (2) control over clarity and framing ($N = 8$). For example, more than half ($N = 6$) wanted to add text to images and configure its color, font, or position, and five mentioned interest in filters and stickers, *"like putting hearts around my dog"* (I7). Two specifically wanted to make memes: *"I got ideas all day long for a really funny little drawing with words under it"* (I10), while a range of other effects were also of interest: blurring (I3), retouching facial appearance (I6), transition effects for videos (I3), and collage making (I7). For clarity and framing, participants desired ways to more accurately capture intended content with good lighting (e.g., scenery (I6), facial expressions (I1, I5)), as well as related post-production steps (e.g., cropping) ($N = 5$).

For digital visual layouts, participants highlighted interest in configuring colors ($N = 9$), font ($N = 7$), theme ($N = 5$), and overall spatial balance ($N = 8$)—this includes presentation slides ($N = 8$), websites/blogs ($N = 5$), visual documents ($N = 2$), resumes ($N = 2$), business cards ($N = 2$), diagrams and charts ($N = 2$), and flyers and posters ($N = 1$). They considered these configurations critical for clarity: *"The layout of text would be as important to a sighted person as it would to a blind person. If the braille is laid out in a weird way, it doesn't make sense to us"* (I9).

Six participants wanted to create illustrations and visual arts, such as logos or icons ($N = 4$), greeting cards ($N = 1$), and cartoons ($N = 1$). I6 had begun to explore potential uses of generative AI tools to create cartoons: *"I would imagine it in my head—some of my old drawings—remember what they looked like, and describe that in a prompt, and see what the tool came up with, and then make some changes."* Five participants wanted to create graphics for tangible products: *"It'd be great to put those drawings on coffee cups."* (I9)

In summary, participants demonstrated interests in evaluating and adjusting *clarity, framing, color, lighting, and overall feeling and style* across different visual media.

4.2 Experiences and Challenges

Here, we present how participants currently engage in digital visual expression, highlighting challenges with inaccessible tools and concerns with sighted help.

4.2.1 Limited to Basic Visual Creation. Overall, participants' visual creation was primarily with simple photo and video tasks ($N = 10$) and formatting limited to text styling ($N = 9$) and theme/template selection ($N = 3$). Most ($N = 8$) had not used expressive edits such as filters or stickers, while the two who tried them with sighted peers felt as if they were *"just pressing buttons and guessing"* (I6). Participants used a range of technology when creating visual content, including: Be My AI[1] ($N = 7$), ChatGPT[2] ($N = 6$), SeeingAI[3] ($N = 4$), iOS native camera with VoiceOver[4] ($N = 3$), Co-Pilot[5] ($N = 2$), OrCam[6] ($N = 1$), Aira[7] ($N = 1$), and Google LookOut[8] ($N = 1$)—mostly used for guiding photo and video taking, evaluating visual outcomes, and generating visual content (e.g., images, front-end code).

Their biggest challenges with configuring visual aesthetics were difficulties checking outcomes ($N = 10$), excessive time and effort due to inaccessible tooling ($N = 10$), and limited visual knowledge ($N = 9$). For example, I5 felt constantly worried about *"how things are aligned, what the camera sees, and how to adjust the camera,"* describing her photography experience as *"daunting,"* while I1 similarly felt a lack of aesthetics-related feedback with presentation slides: *"It's hard to tell the color scheme or if there's crazy lines going everywhere."* Regarding inaccessible edit interactions, participants highlighted difficulties with position-related controls ($N = 7$), such as *"drag and drop"* (I1) and *"point and click"* (I4). In particular, I6 shared frustrations with being expected to visually locate and touch areas in need of editing: *"I would love to retouch my pets' appearance, but you have to get to the exact place in the photo and touch the eyes [for red eye]."* Participants also emphasized on how limited understanding of visual aesthetics hinders their ability to configure visual content, echoing prior work [90, 125]. They desired more guidance for using colors ($N = 9$), fonts ($N = 5$), layouts ($N = 2$), and other visual editing settings ($N = 2$), and for the guidance to be detailed, comprehensive, and updated—*"What is steel blue? I grew up with the 64 pack of Crayola crayons, so everything I referenced goes back to the names of those crayons in the 1970s"* (I6). Further, the guidance should also be actionable (e.g., *"guide you to re-take the picture"* (I8)) and efficient (e.g., avoid *"extra steps to share the photos and switch between apps"* (I2)).

4.2.2 Sighted Help. To create more expressive visual content, participants commonly sought sighted help ($N = 9$). Some preferred handing off the task to sighted helpers with high-level requirements, such as I10: *"I generally give the torch to designers. I tell them, 'This is the concept we want to convey. Please create according to the requirements,'"* while others would try to have a first pass themselves:

[1] https://www.bemyeyes.com/bme-ai/
[2] https://chatgpt.com/
[3] https://www.seeingai.com/
[4] https://www.apple.com/accessibility/vision/
[5] https://copilot.microsoft.com/
[6] http://www.orcam.com/
[7] https://aira.io/
[8] https://play.google.com/store/apps/details?id=com.google.android.apps.accessibility.reveal&pcampaignid=web_share

"The format, size, and everything, I work on them beforehand, and then together we decide on if anything should be changed" (I1).

Participants shared several challenges with sighted help, most commonly a lack of independence ($N = 8$) and communication issues ($N = 8$). Almost all participants desired more independence, as I4 shared: *"I'd really like to be able to do it more on my own."* Some ($N = 2$) were hesitant to ask for help, feeling they might be imposing on others—*"like I'm wasting their time"* (I6), while some ($N = 2$) found it time-consuming to wait for a sighted person to be available. I7 pointed out how visual expressions could be personal or *"an element of grieving"* (I7) and that involving others could feel *"invasive"* (I7): *"I have lots of pictures of my service dog [who passed away], and I'd love to be able to create an album of them on my own, so I can have my emotions and memories depicted the way I'd like to"* (I7). Meanwhile, participants commonly experienced difficulty aligning ideas and a lack of creative agency. Five mentioned having a communication gap: *"It was a little challenging trying to describe to the designer what we wanted [...] they misunderstood what we were trying to convey, and it was hard to tell when they sent mock-ups with images"* (I1). Three experienced a style-mismatch and in turn hoped to make more creative decisions themselves: *"You'd have to match perfectly, or you're just gonna have different aesthetic appeals"* (I3).

Overall, we learned that participants use a combination of technology and sighted help to execute visual expressive tasks, though a range of challenges exist: difficulties checking visual outcomes, concerns over privacy, excessive time and effort, limited visual understanding, communication gaps, and a lack of agency.

4.3 Visual Creative Support Ideation

Participants' ideas for visual creative support initially focused on feedback ($N = 10$) and editing interactions ($N = 7$), though after hearing the audio demos, they became excited about the potential convenience and accessibility of AI-based editing ($N = 10$).

4.3.1 Overall Priorities for Visual Creative Support. First, participants desired more aesthetic-related feedback and guidance ($N = 10$) and comprehensive descriptions for object appearance and composition ($N = 10$). Example aesthetic feedback includes *"how your changes are affecting the image, if it's brighter or darker than normal"* (I1), and whether *"any of the colors [are] standing out or mesh together"* (I8). They also wanted guidance for improving the aesthetics, something that *"just instantly gives you suggestions and tells you things you could change"* (I2), or guidance for matching aesthetics to a certain style: *"if it's a rock poster, it's gonna suggest having these kind of visual elements in it because of the genre of music you're doing"* (I9). Participants requested that feedback and guidance on color, font, and visual effects be comprehensible and actionable without relying on visual concepts ($N = 8$), wanting system designers to *"work closely with people who are blind to understand how these concepts relate to their personal experiences"* (I5).

For editing interactions, participants wanted alternative, accessible ways to perform position-based visual editing, such as visual element and effect placement ($N = 4$), cropping ($N = 2$), and retouching ($N = 1$). For example, I1 envisioned keyboard interactions to place effects and overlays—*"do it incrementally with the arrow keys,"* whereas I7 suggested cropping an image based on object selection—*"have the picture broken down into pieces of visual elements and select which pieces you'd like to remain or remove from the picture."*

4.3.2 Reactions to Audio Demo Ideas. Participants overall felt excited about the audio demos (introduced in Section 3). For the *AI guide* demo, they reacted positively toward its descriptive, aesthetics-related feedback ($N = 10$), actionable and educational edit suggestions ($N = 8$), support for creative agency ($N = 9$), and question and answer functionality ($N = 9$). Participants felt that the AI guide provides information critical for evaluating and planning expressive visual creations while still allowing control by *"suggesting you can do this or you don't have to, so it's not taking over, almost putting the tweak in my hand"* (I3). Many also appreciated the opportunity for visual learning: *"You'll start knowing a couple of things [font and color suggestions] and then that cranks our efficiency up"* (I9).

For the *auto-refiner* demo, participants appreciated its efficiency, convenience ($N = 10$), and provision of inspiration ($N = 6$): *"We can share what we're thinking and see different ideas created by it and then refine it from there, instead of having to have the exact picture or visual from the very beginning"* (I7). Participants considered the auto-refiner a *"grab and go"* (I3) option, especially good for complex and currently inaccessible editing tasks like *"cropping"* (I4, I8), *"positioning graphics"* (I6), and *"aligning and spacing items"* (I10).

Still, participants expressed concerns over AI with important visual expressions, especially related to accuracy ($N = 8$) and authoring effective prompts ($N = 5$). As I2 put it, *"as wonderful as they [AI models] are, you have to say the right combination of words to get what you're looking for"*—which could be challenging, given *"when you don't necessarily know about a certain style or filter, you wouldn't know how to ask for that"* (I2).

Participants thus wanted to use a flexible combination of auto-refiner, guide, and/or sighted help based on specific situations ($N = 10$). As I2 shared: *"If I was doing a presentation in front of the entire faculty, I'd use the AI guide, where I would have a lot more control over all the initial components, and even that, I would still want somebody to look at it."* They emphasized wanting options to iterate on AI-suggested edits—*"I wouldn't want to give up total control. Knowing that I could go in and manually edit it would be nice"* (I4).

4.3.3 Design Implications. Blind visual creators need: (1) *real-time, comprehensive feedback on object appearance, composition, and important aspects of visual aesthetics* (Sections 4.1); (2) *actionable suggestions for improving aesthetics* (Sections 4.2 and 4.3); (3) *options to automatically apply edits based on intent* (Section 4.3); (4) *options to manually edit visuals through accessible interactions, especially with position-based edits* (Sections 4.2 and 4.3); (5) *descriptive labels for editing options without relying on visual experience* (Sections 4.2 and 4.3); and (6) *options to revise and iterate on AI-edits* (Sections 4.2 and 4.3).

5 Design and Implementation of VizXpress

We incorporated the interview implications into an interactive prototype: VizXpress, using it as a design probe to further explore how creative tools could better support the visual expression needs of blind individuals. Powered by large multimodal models [56, 82],

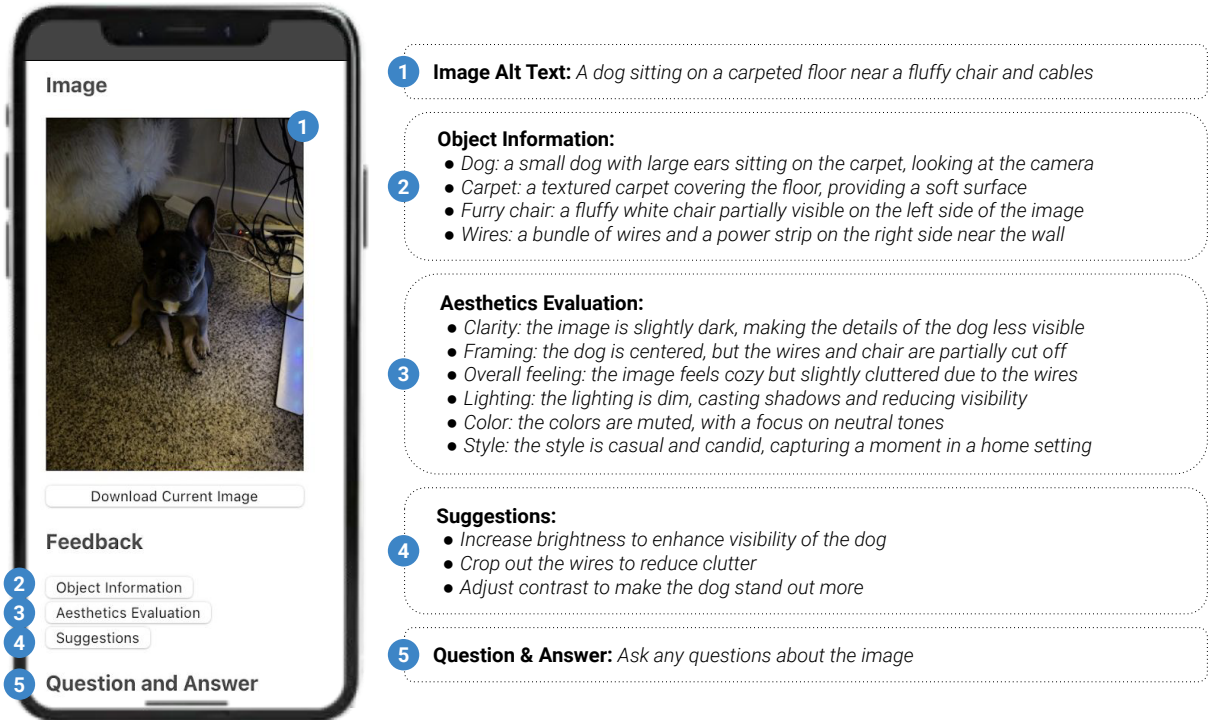


Figure 1: The real-time visual feedback section of VizXpress prototype user interface. The section provides (1) a high-level alt text; (2) object information; (3) aesthetics evaluation; (4) suggestions; (5) question and answer mechanism.

VizXpress provides real-time aesthetics evaluation and guidance, with screen-reader-friendly controls for automated and manual visual edits. To scope our prototype and design probe study, we focused on a commonly desired visual editing task—image editing (Section 4.1).

5.1 Prototype Features

VizXpress was designed as an accessible, one-page web interface, consisting of two sections: a *visual feedback* section (Figure 1) and a *make edits* section (Figure 2).

5.1.1 Visual Feedback. This section provides real-time feedback on the image being edited, including: (1) a *high-level caption*, embedded as alt text for the image; (2) *key objects* extracted from the image, each with a brief description of appearance and position; (3) *aesthetics evaluation* for six aspects of image aesthetics: *clarity*, *framing*, *overall feeling*, *style*, *color*, and *lighting*; and (4) any *visual editing suggestions* applicable to the current image. Additionally, the interface provides a Q&A channel for users to obtain other information about the image from AI. The feedback design was informed by the interview findings (Section 4.3.3) and past image description guidelines (hierarchical description [74, 110] and interactive image exploration [20]).

5.1.2 Editing Features. To balance convenience with creative control, VizXpress supports both *manual* and *automatic* edits:

The research team reviewed common image editing tools (iOS photo editor[9], Adobe Photoshop[10], Instagram photo editing features[11], and Canva photo editing features[12]), identified a set of five basic but expressive visual editing features, and adapted these interactions to be screen-reader friendly:

(1) *Color and Lighting*: Users can incrementally adjust the brightness, saturation, and contrast of an image using "increase," "decrease," and "reset" buttons, with audio feedback that reflects these changes.

(2) *Filters*: Users can apply a small set of filters that we selected by reviewing basic filter options in three common off-the-shelf tools (Canva, iOS image editor, and Instagram): *"Fresco," "Bali," "Nordic," "Chroma," "Aura," "Antiq," "Noir,"* and *"Outrun."* As each tool opts for a different naming scheme, our review was based on the type of visual effect each filter provided, such as black and white or vintage vibes. We then adopted the naming scheme of the off-the-shelf tool (Canva). VizXpress provides a detailed description of each filter, focusing on the perceptual and emotional aspects (e.g., *Bali gives your photo a warm, sunlit feel, with enhanced golden tones*). The full set of filter descriptions can be found in the Supplementary Materials.

(3) *Cropping*: Users can crop by selecting desired focal objects from a list that AI has identified from the original image and selecting how much background padding to include in the cropped area

[9]https://support.apple.com/photos
[10]https://www.adobe.com/products/photoshop
[11]https://www.instagram.com/
[12]https://www.canva.com/

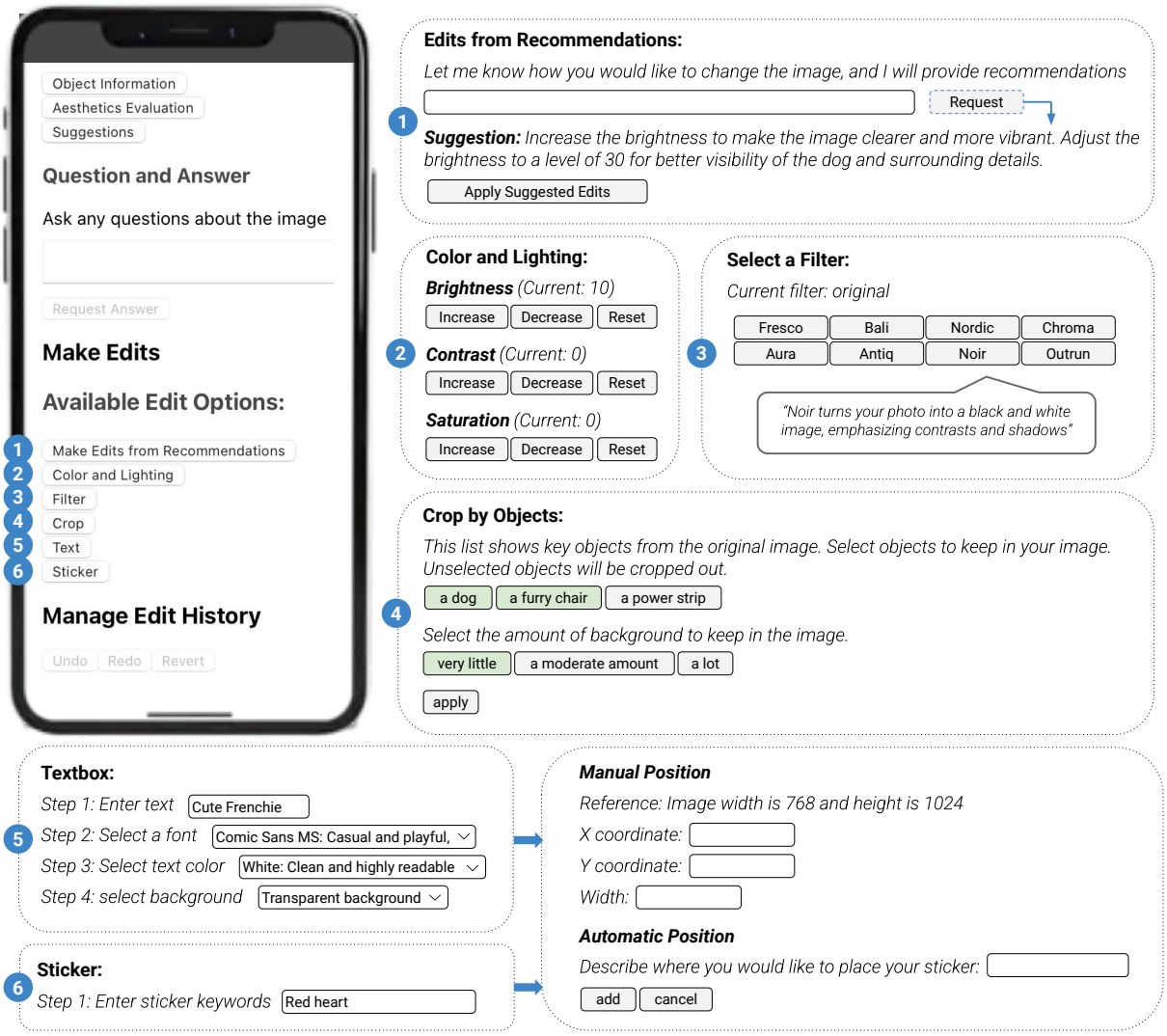


Figure 2: The make edits section of VizXpress prototype user interface. The interface includes automated ((1) *Edits from Recommendations*) and manual (*(2) Color and Lighting, (3) Filter, (4) Crop, (5) Text, and (6) Sticker*) editing options.

(i.e., how tight the crop should be): (1) very little, (2) moderate, or (3) a lot of background. This approach avoids inaccessible interactions that interview participants raised (Section 4.2).

(4) *Text Insertion*: To insert text, the user specifies the text content, styling (size, color, font, background), and position in the image. VizXpress provides two positioning approaches: (a) AI-assisted auto-positioning, where the user describes an ideal insertion location relative to image elements or regions; and (b) manual positioning, where the user specifies the text width and insertion location via (x, y) coordinates.

(5) *Sticker Generation and Placement*: To insert a sticker, the user provides keywords to describe the intended appearance, which the system uses to generate a sticker. Positioning then follows the same design as text insertion, with automatic and manual options.

Users could also *automatically apply edits* by describing how they would like to change the image visually (informed by Section 4.3), based on which the system provides editing recommendations and options to automatically apply them.

5.2 Prototype Implementation

VizXpress was implemented using React.js[13] for the frontend and Python Flask[14] for the backend. We used GPT-4o (gpt-4o-2024-08-06)[15] for all image analysis and generation tasks (conversation history was kept during each image task to provide context). For image manipulation, including color and lighting adjustment, cropping, and text and sticker overlay, we used Fabric.js[16], a Javascript

[13]https://react.dev/
[14]https://flask.palletsprojects.com/en/stable/
[15]https://openai.com/chatgpt/overview/
[16]https://fabricjs.com/

HTML5 canvas library. For automatic edits, the system prompted GPT-4o to generate a set of edit parameters based on user-described intent and manipulated the image with those parameters via Fabric.js.

VizXpress is nearly fully-functional, with the support of two Wizard-of-Oz steps: filter application and connection between an object segmentation script and the web app. First, since there were no readily available presets for applying filters (i.e., predefined image configuration settings), the research team collected the images participants intended to edit before each session and pre-generated filtered versions of them using a common image editing tool (Canva). To make the user experience seem fully functional, when a participant chose to apply a filter, the researcher used an admin panel (a separate web view for the "wizard") to manually upload the corresponding filtered image. The system then automatically reapplied any overlays or edits the participant had already made during that session. Second, due to lack of access to hardware, we needed to run the computer vision model for object-based cropping (Caption-Anything approach [118] to extract key objects from images) on a cloud machine[17] and have the "wizard" quickly upload identified key objects and bounding boxes through the admin panel at the beginning of image tasks.

Below, we briefly describe how each task was prompted (the full list of prompts is in the Supplementary Materials):

- *Visual feedback* prompts request four types of feedback mentioned in Section 5.1, with detail on what to focus on. For example, we prompted for aesthetics evaluations on clarity, framing, message, style, lighting, and color, informed by interview insights (Section 4.1)
- *Question and answer* prompts ask the model to respond to visual questions or offer aesthetic suggestions in concise, accessible language for blind image editors.
- *Edit recommendation and parameter estimation* prompts guide the model to suggest edits aligned with user intent (from six actions in Section 5.1) and provide JSON-formatted parameters for each edit to support automatic image manipulation (e.g., `{"crop": {"objects": ["a keyboard"], 'space": "Moderate"}}`).
- *Overlay location estimation* prompts ask for an appropriate (x, y) coordinate and width for inserting element (stickers or text), based on user intent and object locations from the Caption-Anything output.
- *Sticker generation* prompts request the model to generate a sticker based on user-provided keywords.

5.3 User Scenario

To illustrate interactions supported by VizXpress, we provide a scenario with "Alex," who wants to edit an image of her French bulldog on a carpet (shown as image being edited in Figure 1): Alex begins by examining the image using the *Visual Feedback* panel (Figure 1), learning that it is dimly lit and contains distracting background clutter. Alex first uses a recommended edit (Figure 2) to automatically increase brightness. The aesthetics evaluation suggests that the photo was overexposed, so Alex decides to manually adjust the brightness and contrast. Next, Alex crops the image to eliminate visual clutter by selecting "dog" as the focal object and a *moderate* amount of background around it. She also decides to add the text *"Cute Frenchie"* in a Comic Sans font to match the casual setting and white text to stand out in dark background—the descriptive label helps choosing these styling options. Alex then makes use of the *auto-placement* capability to place this text below the dog's paws. After a final aesthetics check confirming no further suggestions, Alex concludes the image is clear and expressive.

[17] https://colab.research.google.com/

6 Design Probe Study Method: Understanding Information Needs and Tool Design Preferences

We conducted a design probe study using VizXpress to further explore how creative tools should be designed to support expressive visual creation by blind individuals.

6.1 Participants

We recruited 14 participants, including 10 newly recruited through NSITE [80] and the National Federation of the Blind mailing lists, and four returning interview participants who agreed to take part in the design probe study. All participants were 18 years of age or older, identified as blind, primarily used a screen-reader to access technology, and expressed interest in visual expressions. Table 1 presents participants' demographics (D1-D14). Participants were compensated with a $45 gift card for 90 minutes of their time.

6.2 Study Procedure

Sessions were conducted remotely via the Zoom videoconferencing platform. Before the study, participants filled out a survey to provide demographic information, visual editing experience, and one to two images they would like to edit during the study.

Sessions began with the researcher leading a walkthrough of VizXpress using an example image (a simple office desk with a keyboard, mouse, and mug). Participants were encouraged to ask questions and comment on the prototype design. Once participants felt ready, they proceeded to the independent image editing tasks. Participants worked on the images they had uploaded in the pre-study survey and one common image the research team provided (the dog picture in Figure 1). Before participants started to edit their personal images, we asked them to share any knowledge they had about the image, such as purpose and content. We also asked about any specific goals that they hoped to achieve with image edits, and any steps they envisioned performing to achieve those goals. For the common image, we asked participants to imagine the following scenario: *"You met your friend's dog today and really liked him. You wanted to post a cute picture of him on Instagram and want to configure the image to clearly show how cute your friend's dog is. You may also want to add some captions or stickers of your choice."* During the image editing tasks, participants were instructed to freely edit while thinking aloud. The researchers noted editing behaviors, decision making, and any difficulties. Once participants felt satisfied, we then asked about any behavior patterns or comments we had noted. At the end of each image task, we asked about participants' experience and confidence level: *"How would you describe your experience editing this image so far? How confident do you feel that it supports you in achieving your goals, if at all?"* and feedback on specific

features: *"How do you feel about the information this tool provided for the image? How do you feel about the editing functionality you just tried out, in terms of supporting your goals?"*

Once all tasks were completed, we asked participants to reflect on their experience, focusing on the amount of control and expression freedom allowed by the interface design, and any functionality they would like to add, remove, or change. We broadly asked participants to contrast their experience in the study to their typical workflow, and how the tool might or might not fit into that workflow. We ended the study by asking if they were interested in creating other types of visual content beyond editing photos and any prototype adaptation to support those additional creative interests.

6.3 Data Analysis

All sessions were audio-recorded and transcribed. Participants' interactions were logged during the study for analysis, including section navigation and edit actions. Following an inductive thematic analysis approach [16], the first author developed an initial codebook grounded in the transcripts and interaction data, focusing on design insights for an accessible visual creative tool. The first author then independently coded all transcripts. To ensure alignment between data and emerging themes, the third and fourth authors reviewed half of the transcripts (selected by a random number generator). The final themes addressed: (1) the information needs of blind participants in various editing contexts; (2) challenges in applying and reviewing specific edits; and (3) envisioned workflows and use cases for visual expression support.

7 Design Probe Study Findings

We present an overview of participants' image editing tasks, followed by findings on their information needs for expressive visual edits, responses to AI-assisted and manual edits, and envisioned use of visual expression support.

7.1 Image Editing Overview

Participants' personal images included photos of people ($N = 5$), animals ($N = 3$), objects of interest (e.g., flowers, instruments; $N = 3$), scenery ($N = 2$), and graphics ($N = 2$). Common editing goals were improving framing ($N = 8$), enhancing subject focus and appearance ($N = 7$), adjusting color and lighting ($N = 4$), and shifting the mood (e.g., adding humor or drama; $N = 3$). Several ($N = 6$) explored edits without a specific goal, aiming to improve aesthetics. Ten participants used both automatic and manual tools, while four used manual controls only. Editing complexity ranged from simple (e.g., D9 cropped and adjusted lighting) to advanced (e.g., D8, D9, D12 used all features). Table 2 shows examples from the common image task. All participants used the full 90-minute session. D1 and D11 did not have enough time to finish the tasks within the allotted time; the rest focused on cropping ($N = 12$), lighting ($N = 12$), text ($N = 9$), and filters ($N = 5$), with only one using a sticker, though three expressed interest in doing so. Overall, participants were moderately to highly satisfied with their edits ($N = 14$), and most ($N = 10$) felt somewhat confident in their common image results, though some (D2, D10) felt the feedback didn't align with their intended changes, especially around lighting, background cropping, and overlay positioning.

7.2 Information Needs for Expressive Visual Edits

Participants used visual feedback provided by the prototype for checking visual edit results ($N = 14$) and guiding edit decisions ($N = 14$). Often times they were able to achieve their information goals, such as detecting unsatisfactory filter applications: *"it turned my browns into yellows"* (D6) and identifying potential edit ideas: *"it points out things that I, as a blind person, don't really think about"* (D13). When desired information was not covered by the default feedback, participants found the question and answer channel helpful, for checking visual details and obtaining additional suggestions, such as *"which filter would bring out the colors in the clouds?"* (D1) Overall they found the feedback thorough and efficient, providing valuable information about visual aesthetics: *"I really appreciated getting what the tone of the image is, because a lot of times with picture descriptions, they leave a lot of that out, so it fills in a gap"* (D8).

At the same time, participants also shared a range of limitations of the current design of visual feedback, highlighting directions for improvement:

7.2.1 Importance of Matching Visual Feedback to Edit Context. Participants did not find all information useful across editing contexts and desired more edit-specific feedback ($N = 14$): *"There are different things that you're gonna pay more attention to, depending on what your goal is"* (D6). First, the type of visual content impacts participants' information needs. For example, *style* and *overall feeling* were often considered less critical by participants, except for artistic content—such as the book cover images that D2 and D6 worked on: *"I wanted it to convey a romantic style, so for me, style was a big thing"* (D2). Second, participants emphasized different feedback for different editing actions: For *cropping*, object descriptions and framing were the main foci ($N = 14$), as *"it was good to know what was still in the photo object-wise"* (D7). For *text and sticker* application, descriptions to overlay position and appearance were considered relevant ($N = 11$), including absolute position in the image: *"did it [sticker: "cold emoji"] go to the top or the bottom?"* (D14) or position relative to other objects: *"is it [sticker: "pointing finger"] covering anything, and is it pointing in the correct direction?"* (D12) For *color and lighting adjustment*, participants commonly checked the *color*, *lighting*, and *clarity* in the aesthetics evaluation feedback ($N = 14$). However, many found this feedback too *"abstract"* (D14) (see Section 7.2.3).

7.2.2 Importance of Before and After Edit Comparison. Participants ($N = 13$) sometimes had trouble discerning how an image changed from the updated feedback, especially when the effect was subtle (e.g., a small decrease of cropping padding). They wanted visual changes to be verbally highlighted, such as, *"compare what the original image looked like as to what it looks like now"* (D10), or have aspects of the image that changed *"expanded automatically to see exactly how that [image] had changed,"*. D4 also wanted a higher-level summary: *"what did I gain and what did I lose with the filter?"* The feedback should explain why some edits may not substantially change an image, e.g. *"explain that [because the image was taken in low light], it's not gonna get a whole lot better"* (D3). Feedback should also explain why certain edits may lead to unexpected changes in

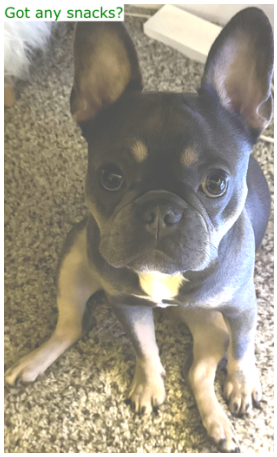
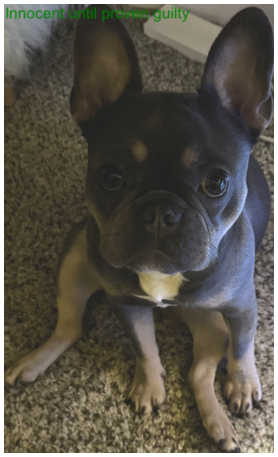
	D3	D7	D8	D10
				
Confidence	*Confident*	*Somewhat confident*	*Somewhat confident*	*Not confident*
Edits	*Crop to focus on dog with little background, increase brightness and contrast, text: "Got any snacks?"*	*Crop to focus on dog with little background, text: "Innocent until proven guilty"*	*Crop to focus on dog and fur with little background, increase brightness and contrast, text: "So Cute!"*	*Crop to focus on dog with moderate background, sticker: "Face with heart shaped eyes"*

Table 2: Example edit result of the common image task from D3, D7, D8, and D10, with each participant's confidence level for the edit result and the specific edits they applied. Confidence levels were qualitatively coded from responses to question: *"How confident do you feel that it supports you in achieving your goals?"* (Section 6).

the description; for example, when an image was cropped to focus on an object, AI tended to pick up more details than before (e.g., specifying the breed of a dog in the common image).

7.2.3 Challenge in Nuanced Color and Lighting Changes. Many participants ($N = 11$) felt the visual feedback often did not help to verify color and lighting adjustment and filter application: *"I was hoping to maybe see a change in the colors or the overall feeling [for filter antiq], but it's still saying that it's clean and modern"* (D12). Participants were unsure whether the lack of change in feedback was a limitation of the AI or because the edits had limited effect on the image. Even with VizXpress interface indicating successful filter application, participants needed more information to make aesthetic editing decisions, which D4 illustrated using a metaphor of comparing the same music piece played on a harp versus piano: *"it's gonna have a different texture and harmonic inflections, [but] you don't know what the Chopin piece on a piano is gonna sound like until you hear it"* (D4). How to best convey subtle lighting and color changes in a non-visual way remains an under-explored question, as D14 pointed out: *"I think that's just inherently the nature of filters and subtle color changes. I'm not sure if there's anything language-wise that would be able to tell me more about it."*

7.2.4 Challenge with Subjective and Filtered Aesthetic Feedback. Some participants ($N = 6$) expressed concerns that AI-generated feedback was overly subjective or filtered, limiting their ability to make independent visual aesthetic decisions, as D12 said, *"Some of them [the aesthetic feedback] are very subjective: overall feeling and style could provide some insights, especially for things more artistic, but I don't necessarily think I'd rely on that as much"* (D12). They felt that for subjective aesthetic decisions, they would prefer to rely on their *"own taste and creativity"* (D11), with D4 further commenting on how it is even difficult to let another person to help with his judgment: *"A sighted person may not have a good answer, either."* The AI feedback was also seen as unhelpful for some sensitive topics. D12, for example, was not able to get useful feedback about blemishes on his face in an edited image:

> *"Interesting [that] it's not able to provide more details about it [...] I've noticed some AI feedback always comments on things in a really positive way and never says, you know, this person or this object doesn't look very put together, so that, to me, is a problem. That's bias, because that's access to the same information that everybody else would be able to see. I would be cautious about the feedback that it gives knowing about the biases, especially when it tries to paint everything rosy and nice."* (D12)

In turn, participants wanted more feedback that is *"objective"*(D12), *"factual"*(D10), and *"verifiable by consensus"* (D12).

7.3 Reaction to AI-Assisted vs. Manual Edits

Participants found the AI-assisted and manual edit functions to be useful and complementary, and provided ideas for further improvement.

7.3.1 Reaction to Manual Edits. Participants were excited about the accessibility of our manual edit design and the control it provided ($N = 13$). Many ($N = 12$) found object-based cropping accessible, describing it as *"intuitive"* (D4) and *"comfortable to use"* (D6, D11). Participants also especially liked the informative labels for unfamiliar visual editing options ($N = 10$), such as filters, fonts, and font colors. As D13 shared:

> *"The thing that really made me happy was the way you guys described the fonts because a lot of blind people don't know what certain fonts look like or what they do. I also like how you described the colors like how you referred to orange as like a kind of playful color."*(D13)

Participants felt that more control and precision ($N = 10$) and guidance on position-based editing ($N = 11$) would further support their expression. Specifically, they wanted more precision in selecting what to crop (e.g., specifying background padding beyond the current three levels, options to remove objects or a part of an object, cropping by coordinates) and in placing text and stickers—such as by *"automatically centering horizontal or vertical"* (D6) and *"moving up or down by 10 pixels [each time]"* (D10). D4 expressed the overall desire for more equality in visual expression control: *"How many options [to crop] would a sighted person have? Would it just be three? I want to have exactly what you have."* At the same time, more position-related guidance is necessary for precise control, such as instructions for coordinates ($N = 4$)—*"where does* $(0, 0)$ *start?"* (D12)—and aesthetic placement: *"I'm not sure where those things [stickers] are typically put on a photo, so that could be something to add as a suggestion"* (D5). More information about object positioning (e.g., *"the* (x, y) *coordinates for image items"* (D8)) could also be helpful, as well as a description to the effect of different cropping options, such as *"how much background each option leaves in the image and have a really short feedback to say this is what the image should look like with these options, like a text version of a preview"* (D12).

7.3.2 Reaction to AI-Assisted Edits. Overall, participants considered the AI-assisted edits to be useful ($N = 14$) and especially good when they were in a time crunch or in need of inspiration. D11, for example, commented on the inspiration aspect: *"I can learn what potential edits there are that I might not have considered."* D1 also appreciated the efficiency of not having to *"go back into settings and figure out what those kinds of parameters are."* D6 felt that AI could sometimes provide more holistic feedback compared to sighted peers: *"AI is not one subjective opinion, but an aggregated one, it's more like I've asked multiple friends."* Participants thus preferred AI-assisted edits over the manual ones for tasks that they were unfamiliar with, as D7 shared: *"I thought it would have a better feel for the brightness and contrast than I would."*

As with any AI-assisted tools, participants encountered various accuracy issues, such as ineffective and inconsistent edit recommendations (e.g., *"keeps going back and forth [on brightness]"* (D13)) and inaccurate feedback (e.g., mistakenly describing object parts being cropped out). Participants were aware of the potential accuracy risks and alerted to them: *"It could sound beautiful with AI, and then somebody looks at it and thinks that looks awful"* (D1). As a result, participants wanted to know when a mistake happened and have the choice to undo it: *"Like collaborating with a good smart friend—I don't have to agree with everything they say. I'm definitely going to take it into consideration, but you say, go brighter and brighter, like, sorry, buddy, you're wrong. We're gonna go back"* (D6). Participants felt that detailed feedback (D2), a report on confidence level (D9, D12), the ability to review and manipulate specific steps within the automated edits (D12), and improvement on AI accuracy over time (D3, D13) will help them feel more confident with AI-assisted edits.

Overall, participants noted pros and cons for both AI-assisted and manual edits: *"I like the speed and easiness of the AI, but then I also don't get control, so, it's kind of a trade off, depending on the situation"* (D6). Manual editing was deemed as more valuable when the image was important, if the AI could not produce the intended result, and when the participant had a clear idea of the intended visual edit; on the other hand, automated editing was seen as more efficient and fast, when there was a high confidence in the AI being able to make the particular edit. Many participants shared that they would still seek verification from a human or another AI even when using the two tools in combination.

7.4 Envisioned Use of Visual Expression Support

All participants were excited about the potential of visual expression support ($N = 14$), as D5 shared:

> *"I am fully blind and born blind, but I have kind of a good imagination, so I see them in my head. I feel like this is what I've been waiting for. I'm a travel advisor among other things, and I feel like I would spend more time designing things for social media because I feel like I could, and I would know whether it looked visually appealing or not. It's gonna open a lot of work opportunities and new skills for people who are blind."* (D5)

Participants valued the independence provided by AI-based feedback and editing support as well as the opportunity to own their visual expressions ($N = 11$)—D8 contrasted the experience of using the prototype with her usual workflow of asking a sighted friend or family member for editing help: *"This way, I feel like I have more freedom to edit the image how I want it to be and not necessarily what someone else thinks looks good"* (D8). They also felt that learning and exploring about visual expressions this way helps them better understand their preferences ($N = 11$): *"If there are tools like this available, then people can get more used to dealing with images, and it (non-visual image editing) will become more mainstream"* (D10). Participants wanted to use AI-based feedback and edit support for personal or professional branding (e.g., book graphics (D2), travel agency (D5), online store (D6)), document and presentation materials (e.g., adding images (D1), choosing fonts (D10)), photos and videos for professional events (D7, D12) or social media—*"put up the captions"* (D3), *"touching up pictures of pets"* (D7)—and various artistic explorations. In envisioning incorporating this type of support into their workflows, participants ($N = 13$) commonly found it useful for *"getting these images started and communicating what my vision is"* (D6). However, for completing the end-product, all needed more time to test the tool with sighted helper ($N = 14$), a different AI ($N = 5$), or tactile-based methods (e.g., dynamic tactile displays, embossed braille) ($N = 3$).

We also observed that visual edits related to color and lighting adjustments ($N = 14$), text overlay ($N = 14$), cropping ($N = 14$), and filter application ($N = 13$) were particularly popular among participants. In contrast, stickers received mixed reactions—nine participants appreciated the feature, while five found it less appealing—for instance, D2 remarked: *"I don't see myself wanting to use stickers... It just isn't my cup of tea."* Beyond the visual editing capabilities supported by the prototype, participants shared interest in applying this type of AI-support for other visual expressive tasks, including generative visual edits ($N = 7$) (e.g., editing AI-generated imagery, making graphic modifications within existing images), guided photo- and video- taking ($N = 5$), video editing ($N = 5$), designing presentation slides and flyers ($N = 5$), and exploring more style and effect options ($N = 4$). Participants commented on how the prototype's feedback could inform tasks such as camera framing—having the AI *"calibrated in a way that it would tell me the picture has a part of my shoe or some of the wall"* (D4), *"lighting and filtering"* tasks for videos (D3), and visual design tasks like creating a pamphlet or flyer (D12).

8 Discussion

Through an in-depth interview and design probe study, we uncovered a wide range of creative interests (RQ1) (e.g., aesthetic refinement, visual effect exploration, styling configuration), motivations (RQ1) (e.g., personal and professional branding, artistic pursuit), and challenges (RQ2) (e.g., limited creative agency, difficulty evaluating aesthetics) among blind individuals who are interested in visual creation. We identified key design insights critical to supporting their expressive freedom (RQ3), including context-specific aesthetic feedback, accessible editing interactions that combine automation with manual control, and opportunities for ongoing visual learning. Although participants were excited by the increased autonomy in expressive visual creation enabled by our design explorations, much work remains to fully support blind individuals in freely expressing themselves through this media. Here, we discuss how our findings extend prior knowledge in accessible creativity support and outline directions for future research.

8.1 Support Subtle Visual Feedback for More Expressive Freedom

The design probe study confirms the effectiveness of some established image description guidelines in the context of expressive visual editing, particularly regarding context-specific informational needs [108], hierarchical structuring of visual information [74, 110, 128], and verification loops [20, 48, 124]. However, findings revealed unique challenges associated with perceiving and evaluating subtle visual changes—especially those involving shifts in lighting, color, or stylistic "vibe." While current AI models produce generally useful descriptions, they fall short in effectively communicating these subtle yet critical aesthetic nuances, causing confusion during the evaluation of visual edits. An essential avenue for future research lies in defining objective yet sufficiently descriptive feedback for evaluating non-visual aesthetics, a challenge also recognized in other visual description contexts [21]. This research direction could benefit from insights in the art criticism and visual design literature (e.g., [5, 32, 55]). Clear "before-and-after" comparisons could also help convey subtle changes, a feature seemingly within the capability of existing large multimodal models (e.g., [82]), though systematic evaluations are necessary to determine their actual effectiveness for blind users. Integrating tactile and braille feedback presents another promising dimension of sensory engagement to complement AI-generated descriptions, echoing prior work [62, 127]. Nevertheless, accurately translating visual aesthetics—such as framing and nuanced color representation—into tactile formats requires substantial experimentation and refinement.

8.2 Objective and Critical Evaluation from AI

An important concern flagged by our participants relates to AI's reluctance or inability to provide candid evaluations on potentially sensitive or subjective visual matters, possibly due to privacy or safety constraints [34, 72]. Participants also criticized the overly positive or superficial nature of AI-generated feedback, particularly in contexts involving personal or potentially critical aesthetics (e.g., facial appearance, cleanliness, fashion). This critique echoes frustrations within the blind community regarding insufficient detail in AI-generated image descriptions, limiting their ability to fully access visual information [89]. Without transparent, objective, and tactful feedback, blind creators risk misunderstanding or misrepresenting their intended visual communication. Addressing this challenge calls for research into models capable of offering straightforward yet sensitive evaluations, potentially drawing insights from literature related to representation in image description or critique methods used in artistic communities [6, 12, 38]. Overcoming this delicate yet critical limitation is essential to empowering blind creators to confidently engage in authentic visual expression, especially in contexts involving self-image or identity.

8.3 Enhanced Non-Visual Edit Control through Interaction Design

Several existing techniques for non-visual editing—such as natural language commands [20, 48], visual understanding support [90], and object-based cropping or obfuscation [124]—proved applicable to aesthetic configuration. However, our study highlights challenges with natural language interfaces when users lack a clear visual aesthetic goal, underscoring the importance of systems that can recommend or guide visual edits. We found early promise in offering styling instructions framed through perception (e.g., shape or size), mood (e.g., playful or formal), and utility (e.g., text clarity against background). This descriptive framing helped scaffold choices around elements like font, color, and filters, though additional work is needed to extend this approach to other domains where genre-specific context may be important—for instance, meme creation. More precise spatial control remains a key area for improvement. Tasks like cropping, overlay positioning, and retouching require fine-grained adjustments beyond what current systems typically offer. While object- and region-based referencing provides a helpful baseline, future tools should also support refined manipulation, such as nudging, resizing, or removing elements, coupled with a quick verbal preview for each option. Tactile input could be another promising direction to support spatial precision.

8.4 Aesthetic and Expressive Creation Beyond Image Editing

Our participants demonstrated expressive interests across a wide range of visual media and envisioned applying functionalities of VizXpress to video editing, generative art, and visual layout design. These tasks each introduce unique accessibility challenges and design complexities that demand targeted research. For example, although aesthetic configuration for videos might directly benefit from insights gained around filters and overlays, the temporal dimension will likely require fundamentally new interaction techniques and feedback models, where insights from [49] could help. Similarly, generative tasks (e.g., logo creation) may require enhanced communication of abstract visual qualities [21]. At the same time, many blind individuals prefer to engage with aesthetic creation through non-visual modalities such as audio or tactile formats. Future research should examine how to support more inclusive content sharing on visually dominant platforms. One direction could be to encourage multimodal content creation and sharing, such as tools for editing and posting aesthetic audio, optionally paired with auto-generated graphical representations for visual audiences. Such approaches could expand both social participation and creative agency for blind users across a wider range of visual culture.

8.5 Continuous Visual Learning

Expressive visual engagement among blind individuals could evolve with changing preferences, experiences, and cultural exposure. Participants emphasized the importance of ongoing opportunities to build visual literacy within creative tools themselves. Future systems should support continuous learning by embedding interactive tutorials, context-aware aesthetic suggestions, or prompts that encourage visual exploration tailored to a user's evolving interests. Should accessible visual creation become more integrated into mainstream platforms or workflows, research should remain attentive to how blind creators' perspectives and aesthetic priorities evolve.

8.6 Limitations

Our prototype evaluations occurred in controlled sessions rather than through extended field use, limiting insights into long-term adoption and everyday integration. Future research should include longitudinal studies to assess real-world usability and sustained engagement. Additionally, our team's positionality—sighted researchers with moderate screen-reader experience—may have influenced our interpretations and design priorities.

9 Conclusion

In conclusion, this study highlights the potential of AI support to improve the accessibility of visual aesthetic expression. By understanding and designing for blind creators' visual expression needs and accessibility barriers, our research highlighted the importance of context-specific aesthetic feedback and accessible editing controls that balances automation and creative agency. The VizXpress design provides a concrete illustration of how to enable accessible visual creativity, offering a starting point that informs future designs. Ultimately, supporting visual content creation is not only about accessibility but also about empowering free participation in the creative culture.

Acknowledgments

We thank our participants for their thoughtful insights. This work was partially supported by Apple Inc and the University of Washington CREATE Center.

Conflict of interest disclosure: Leah Findlater is also employed by and has a conflict of interest with 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] Dustin Adams, Sri Kurniawan, Cynthia Herrera, Veronica Kang, and Natalie Friedman. 2016. Blind photographers and VizSnap: A long-term study. In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility*. 201–208.

[2] 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. doi:10.1145/2504335.2504360

[3] Rudaiba Adnin and Maitraye Das. 2024. "I look at it as the king of knowledge": How Blind People Use and Understand Generative AI Tools. In *Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility* (St. John's, NL, Canada) *(ASSETS '24)*. Association for Computing Machinery, New York, NY, USA, Article 64, 14 pages. doi:10.1145/3663548.3675631

[4] Dragan Ahmetovic, Nahyun Kwon, Uran Oh, Cristian Bernareggi, and Sergio Mascetti. 2021. Touch Screen Exploration of Visual Artwork for Blind People. In *Proceedings of the Web Conference 2021* (Ljubljana, Slovenia) *(WWW '21)*. Association for Computing Machinery, New York, NY, USA, 2781–2791. doi:10.1145/3442381.3449871

[5] Duaa Alashari. 2021. The significance of Feldman method in art criticism and art education. *International Journal of Psychosocial Rehabilitation* 25, 2 (2021), 877–884.

[6] Rahaf Alharbi, Pa Lor, Jaylin Herskovitz, Sarita Schoenebeck, and Robin N. Brewer. 2024. Misfitting With AI: How Blind People Verify and Contest AI Errors. In *Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility* (St. John's, NL, Canada) *(ASSETS '24)*. Association for Computing Machinery, New York, NY, USA, Article 61, 17 pages. doi:10.1145/3663548.3675659

[7] Saki Asakawa, João Guerreiro, Daisuke Sato, Hironobu Takagi, Dragan Ahmetovic, Desi Gonzalez, Kris M Kitani, and Chieko Asakawa. 2019. An independent and interactive museum experience for blind people. In *Proceedings of the 16th International Web for All Conference*. 1–9.

[8] Elisabeth Salzhauer Axel and Nina Sobol Levent. 2003. *Art beyond sight: a resource guide to art, creativity, and visual impairment*. American Foundation for the Blind.

[9] Jan Balata, Zdenek Mikovec, and Lukas Neoproud. 2015. BlindCamera: Central and Golden-ratio Composition for Blind Photographers. In *Proceedings of the Mulitimedia, Interaction, Design and Innnovation* (Warsaw, Poland) *(MIDI '15)*. Association for Computing Machinery, New York, NY, USA, Article 8, 8 pages. doi:10.1145/2814464.2814472

[10] Aadit Barua, Karim Benharrak, Meng Chen, Mina Huh, and Amy Pavel. 2025. Lotus: Creating Short Videos From Long Videos With Abstractive and Extractive Summarization. In *Proceedings of the 30th International Conference on Intelligent User Interfaces (IUI '25)*. Association for Computing Machinery, New York, NY, USA, 967–981. doi:10.1145/3708359.3712090

[11] 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. doi:10.1145/3173574.3173650

[12] Cynthia L. Bennett, Cole Gleason, Morgan Klaus Scheuerman, Jeffrey P. Bigham, Anhong Guo, and Alexandra To. 2021. "It's Complicated": Negotiating Accessibility and (Mis)Representation in Image Descriptions of Race, Gender, and Disability. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 375, 19 pages. doi:10.1145/3411764.3445498

[13] Cynthia L Bennett, Renee Shelby, Negar Rostamzadeh, and Shaun K Kane. 2024. Painting with Cameras and Drawing with Text: AI Use in Accessible Creativity. In *Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility* (St. John's, NL, Canada) *(ASSETS '24)*. Association for Computing Machinery, New York, NY, USA, Article 5, 19 pages. doi:10.1145/3663548.3675644

[14] Jeffrey P. Bigham, Chandrika Jayant, Hanjie Ji, Greg Little, Andrew Miller, Robert C. Miller, Robin Miller, Aubrey Tatarowicz, Brandyn White, Samual White, and Tom Yeh. 2010. VizWiz: nearly real-time answers to visual questions. In *Proceedings of the 23nd Annual ACM Symposium on User Interface Software and Technology* (New York, New York, USA) *(UIST '10)*. Association for Computing Machinery, New York, NY, USA, 333–342. doi:10.1145/1866029.1866080

[15] Jens Bornschein and Gerhard Weber. 2017. Digital Drawing Tools for Blind Users: A State-of-the-Art and Requirement Analysis. In *Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments* (Island of Rhodes, Greece) *(PETRA '17)*. Association for Computing Machinery, New York, NY, USA, 21–28. doi:10.1145/3056540.3056542

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

[17] Tim Brooks, Aleksander Holynski, and Alexei A Efros. 2023. Instructpix2pix: Learning to follow image editing instructions. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 18392–18402.

[18] Amanda Cachia. 2013. 'disabling'the museum: Curator as infrastructural activist. *Journal of Visual Art Practice* 12, 3 (2013), 257–289.

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

[20] Ruei-Che Chang, Yuxuan Liu, Lotus Zhang, and Anhong Guo. 2024. EditScribe: Non-Visual Image Editing with Natural Language Verification Loops. In *Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility* (St. John's, NL, Canada) *(ASSETS '24)*. Association for Computing Machinery, New York, NY, USA, Article 65, 19 pages. doi:10.1145/3663548.3675599

[21] Arnavi Chheda-Kothary, Ritesh Kanchi, Chris Sanders, Kevin Xiao, Aditya Sengupta, Melanie Kneitmix, Jacob O. Wobbrock, and Jon E. Froehlich. 2025. ArtInsight: Enabling AI-Powered Artwork Engagement for Mixed Visual-Ability Families. In *Proceedings of the 30th International Conference on Intelligent User Interfaces (IUI '25)*. Association for Computing Machinery, New York, NY, USA, 190–210. doi:10.1145/3708359.3712082

[22] Gina Clepper, Emma J. McDonnell, Leah Findlater, and Nadya Peek. 2025. "What Would I Want to Make? Probably Everything": Practices and Speculations of Blind and Low Vision Tactile Graphics Creators. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 1159, 16 pages. doi:10.1145/3706598.3714173

[23] Chris Creed. 2018. Assistive technology for disabled visual artists: exploring the impact of digital technologies on artistic practice. *Disability & Society* 33, 7 (2018), 1103–1119.

[24] Hai Dang, Frederik Brudy, George Fitzmaurice, and Fraser Anderson. 2023. Worldsmith: Iterative and expressive prompting for world building with a generative ai. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–17.

[25] Maitraye Das, Alexander J. Fiannaca, Meredith Ringel Morris, Shaun K. Kane, and Cynthia L. Bennett. 2024. From Provenance to Aberrations: Image Creator and Screen Reader User Perspectives on Alt Text for AI-Generated Images. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 900, 21 pages. doi:10.1145/3613904.3642325

[26] DIAGRAM Center. 2017. *Specific Guidelines: Art, Photos & Cartoons.* DIAGRAM Center. http://diagramcenter.org/specific-guidelines-final-draft.html Accessed: 2025-04-15.

[27] Alice Drew and Salvador Soto-Faraco. 2024. Perceptual oddities: assessing the relationship between film editing and prediction processes. *Philosophical Transactions of the Royal Society B* 379, 1895 (2024), 20220426.

[28] Peitong Duan, Jeremy Warner, Yang Li, and Bjoern Hartmann. 2024. Generating Automatic Feedback on UI Mockups with Large Language Models. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 6, 20 pages. doi:10.1145/3613904.3642782

[29] Eric Eggert and Shadi Abou-Zahra. 2024. *Images Tutorial.* World Wide Web Consortium (W3C). https://www.w3.org/WAI/tutorials/images/ Accessed: 2025-04-15.

[30] Tsu-Jui Fu, William Yang Wang, Daniel McDuff, and Yale Song. 2022. Doc2ppt: Automatic presentation slides generation from scientific documents. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 36. 634–642.

[31] Sanjana Gautam, Pranav Narayanan Venkit, and Sourojit Ghosh. 2024. From Melting Pots to Misrepresentations: Exploring Harms in Generative AI. *arXiv preprint arXiv:2403.10776* (16 3 2024). doi:10.48550/arXiv.2403.10776 Accessed: 2025-04-15.

[32] Jeremy Glatstein. 2009. Formal visual analysis: The elements & principles of composition. *The Kennedy* (2009).

[33] 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. doi:10.1145/3517428.3550372

[34] Google Cloud. 2025. *Responsible AI and usage guidelines for Imagen.* https://cloud.google.com/vertex-ai/generative-ai/docs/image/responsible-ai-imagen Accessed: 2025-04-15.

[35] William Grussenmeyer and Eelke Folmer. 2016. AudioDraw: user preferences in non-visual diagram drawing for touchscreens. In *Proceedings of the 13th International Web for All Conference* (Montreal, Canada) *(W4A '16)*. Association for Computing Machinery, New York, NY, USA, Article 22, 8 pages. doi:10.1145/2899475.2899483

[36] Ananya Gubbi Mohanbabu and Amy Pavel. 2024. Context-Aware Image Descriptions for Web Accessibility. In *Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility* (St. John's, NL, Canada) *(ASSETS '24)*. Association for Computing Machinery, New York, NY, USA, Article 62, 17 pages. doi:10.1145/3663548.3675658

[37] Kozue Handa, Hitoshi Dairoku, and Yoshiko Toriyama. 2010. Investigation of priority needs in terms of museum service accessibility for visually impaired visitors. *British journal of visual impairment* 28, 3 (2010), 221–234.

[38] 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 (AIES '21)*. ACM, 543–554. doi:10.1145/3461702.3462620

[39] 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. doi:10.1145/2470654.2481292

[40] Susumu Harada, Jacob O. Wobbrock, and James A. Landay. 2007. Voicedraw: a hands-free voice-driven drawing application for people with motor impairments. In *Proceedings of the 9th International ACM SIGACCESS Conference on Computers and Accessibility* (Tempe, Arizona, USA) *(Assets '07)*. Association for Computing Machinery, New York, NY, USA, 27–34. doi:10.1145/1296843.1296850

[41] Simon Hayhoe. 2013. Expanding our vision of museum education and perception: An analysis of three case studies of independent blind arts learners. *Harvard Educational Review* 83, 1 (2013), 67–86.

[42] Raquel Hervás, Alberto Díaz, Matías Amor, Alberto Chaves, and Víctor Ruiz. 2024. Self-guided Spatial Composition as an Additional Layer of Information to Enhance Accessibility of Images for Blind Users. In *Proceedings of the XXIV International Conference on Human Computer Interaction* (A Coruña, Spain) *(Interacción '24)*. Association for Computing Machinery, New York, NY, USA, Article 4, 8 pages. doi:10.1145/3657242.3658590

[43] Naoki Hirabayashi, Masakazu Iwamura, Zheng Cheng, Kazunori Minatani, and Koichi Kise. 2023. VisPhoto: Photography for People with Visual Impairments via Post-Production of Omnidirectional Camera Imaging. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility* (New York, NY, USA) *(ASSETS '23)*. Association for Computing Machinery, New York, NY, USA, Article 6, 17 pages. doi:10.1145/3597638.3608422

[44] Jonggi Hong and Hernisa Kacorri. 2024. Understanding How Blind Users Handle Object Recognition Errors: Strategies and Challenges. In *The 26th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '24)*. ACM, 1–15. doi:10.1145/3663548.3675635

[45] Qirui Huang, Min Lu, Joel Lanir, Dani Lischinski, Daniel Cohen-Or, and Hui Huang. 2024. Graphimind: Llm-centric interface for information graphics design. *arXiv preprint arXiv:2401.13245* (2024).

[46] Mina Huh, YunJung Lee, Dasom Choi, Haesoo Kim, Uran Oh, and Juho Kim. 2022. Cocomix: Utilizing Comments to Improve Non-Visual Webtoon Accessibility. 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 607, 18 pages. doi:10.1145/3491102.3502081

[47] Mina Huh and Amy Pavel. 2024. DesignChecker: Visual Design Support for Blind and Low Vision Web Developers. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology* (Pittsburgh, PA, USA) *(UIST '24)*. Association for Computing Machinery, New York, NY, USA, Article 142, 19 pages. doi:10.1145/3654777.3676369

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

[49] Mina Huh, Saelyne Yang, Yi-Hao Peng, Xiang 'Anthony' Chen, Young-Ho Kim, and Amy Pavel. 2023. AVscript: Accessible Video Editing with Audio-Visual

Scripts. 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 796, 17 pages. doi:10.1145/3544548.3581494
[50] Imagine.art. 2025. *Imagine.art.* https://www.imagine.art/ Accessed: 2025-04-15.
[51] Adobe Inc. 2025. *Adobe Firefly.* https://www.adobe.com/products/firefly.html Accessed: 2025-04-15.
[52] Chandrika Jayant, Hanjie Ji, Samuel White, and Jeffrey P. Bigham. 2011. Supporting blind photography. In *The Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility* (Dundee, Scotland, UK) *(ASSETS '11)*. Association for Computing Machinery, New York, NY, USA, 203–210. doi:10.1145/2049536.2049573
[53] Lucy Jiang, Crescentia Jung, Mahika Phutane, Abigale Stangl, and Shiri Azenkot. 2024. "It's Kind of Context Dependent": Understanding Blind and Low Vision People's Video Accessibility Preferences Across Viewing Scenarios. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 897, 20 pages. doi:10.1145/3613904.3642238
[54] Hesham M. Kamel and James A. Landay. 2000. A study of blind drawing practice: creating graphical information without the visual channel. In *Proceedings of the Fourth International ACM Conference on Assistive Technologies* (Arlington, Virginia, USA) *(Assets '00)*. Association for Computing Machinery, New York, NY, USA, 34–41. doi:10.1145/354324.354334
[55] Keith Kenney and Linda M Scott. 2003. A review of the visual rhetoric literature. *Persuasive imagery* (2003), 17–56.
[56] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. 2023. Segment anything. In *Proceedings of the IEEE/CVF international conference on computer vision*. 4015–4026.
[57] Georgina Kleege. 2017. *More than meets the eye: What blindness brings to art.* Oxford University Press.
[58] Chinmay Kulkarni, Stefania Druga, Minsuk Chang, Alex Fiannaca, Carrie Cai, and Michael Terry. 2023. A word is worth a thousand pictures: Prompts as AI design material. *arXiv preprint arXiv:2303.12647* (2023).
[59] Martin Kurze. 1996. TDraw: a computer-based tactile drawing tool for blind people. In *Proceedings of the Second Annual ACM Conference on Assistive Technologies* (Vancouver, British Columbia, Canada) *(Assets '96)*. Association for Computing Machinery, New York, NY, USA, 131–138. doi:10.1145/228347.228368
[60] Jaewook Lee, Yi-Hao Peng, Jaylin Herskovitz, and Anhong Guo. 2021. Image Explorer: Multi-Layered Touch Exploration to Make Images Accessible. 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 69, 4 pages. doi:10.1145/3441852.3476548
[61] Kyungjun Lee, Jonggi Hong, Simone Pimento, Ebrima Jarjue, and Hernisa Kacorri. 2019. Revisiting Blind Photography in the Context of Teachable Object Recognizers. In *Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility* (Pittsburgh, PA, USA) *(ASSETS '19)*. Association for Computing Machinery, New York, NY, USA, 83–95. doi:10.1145/3308561.3353799
[62] Seonghee Lee, Maho Kohga, Steve Landau, Sile O'Modhrain, and Hari Subramonyam. 2024. AltCanvas: A Tile-Based Editor for Visual Content Creation with Generative AI for Blind or Visually Impaired People. In *Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility* (St. John's, NL, Canada) *(ASSETS '24)*. Association for Computing Machinery, New York, NY, USA, Article 70, 22 pages. doi:10.1145/3663548.3675600
[63] Franklin Mingzhe Li, Franchesca Spektor, Meng Xia, Mina Huh, Peter Cederberg, Yuqi Gong, Kristen Shinohara, and Patrick Carrington. 2022. "It Feels Like Taking a Gamble": Exploring Perceptions, Practices, and Challenges of Using Makeup and Cosmetics for People with Visual Impairments. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 266, 15 pages. doi:10.1145/3491102.3517490
[64] 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. doi:10.1145/3544548.3580941
[65] 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. doi:10.1145/3290605.3300436
[66] Junchen Li, Garreth W. Tigwell, and Kristen Shinohara. 2021. Accessibility of High-Fidelity Prototyping Tools. 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 493, 17 pages. doi:10.1145/3411764.3445520
[67] Jiasheng Li, Zeyu Yan, Ebrima Haddy Jarjue, Ashrith Shetty, and Huaishu Peng. 2022. TangibleGrid: Tangible Web Layout Design for Blind Users. 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 47, 12 pages. doi:10.1145/3526113.3545627
[68] Jongho Lim, Yongjae Yoo, Hanseul Cho, and Seungmoon Choi. 2019. TouchPhoto: Enabling Independent Picture Taking and Understanding for Visually-Impaired Users. In *2019 International Conference on Multimodal Interaction* (Suzhou, China) *(ICMI '19)*. Association for Computing Machinery, New York, NY, USA, 124–134. doi:10.1145/3340555.3353728
[69] Gitte Lindgaard. 2007. Aesthetics, Visual Appeal, Usability, and User Satisfaction. *Australian journal of emerging technologies and society* (2007).
[70] Wenhao Y. Luebs, Garreth W. Tigwell, and Kristen Shinohara. 2024. Understanding Expert Crafting Practices of Blind and Low Vision Creatives. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '24)*. Association for Computing Machinery, New York, NY, USA, Article 346, 8 pages. doi:10.1145/3613905.3650960
[71] Kelly Avery Mack, Rida Qadri, Remi Denton, Shaun K. Kane, and Cynthia L. Bennett. 2024. "They only care to show us the wheelchair": disability representation in text-to-image AI models. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 288, 23 pages. doi:10.1145/3613904.3642166
[72] Meta. 2021. *Using AI to Improve Photo Descriptions for People Who Are Blind and Visually Impaired.* https://about.fb.com/news/2021/01/using-ai-to-improve-photo-descriptions-for-blind-and-visually-impaired-people/ Accessed: 2025-04-15.
[73] Jeff Mitscherling and Paul Fairfield. 2019. *Artistic Creation: A Phenomenological Account.* Rowman & Littlefield.
[74] 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. doi:10.1145/3173574.3173633
[75] Meredith Ringel Morris, Annuska Zolyomi, Catherine Yao, Sina Bahram, Jeffrey P. Bigham, and Shaun K. Kane. 2016. "With most of it being pictures now, I rarely use it": Understanding Twitter's Evolving Accessibility to Blind Users. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 5506–5516. doi:10.1145/2858036.2858116
[76] Gillian M Morriss-Kay. 2010. The evolution of human artistic creativity. *Journal of anatomy* 216, 2 (2010), 158–176.
[77] Vishnu Nair, Hanxiu 'Hazel' Zhu, and Brian A. Smith. 2023. ImageAssist: Tools for Enhancing Touchscreen-Based Image Exploration Systems for Blind and Low Vision Users. 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 76, 17 pages. doi:10.1145/3544548.3581302
[78] Zheng Ning, Brianna L Wimer, Kaiwen Jiang, Keyi Chen, Jerrick Ban, Yapeng Tian, Yuhang Zhao, and Toby Jia-Jun Li. 2024. SPICA: Interactive Video Content Exploration through Augmented Audio Descriptions for Blind or Low-Vision Viewers. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 902, 18 pages. doi:10.1145/3613904.3642632
[79] Don Norman. 2007. *Emotional design: Why we love (or hate) everyday things.* Basic books.
[80] NSITE. 2025. NSITE - A Vision for Talent. https://nsite.org/ Accessed: April 17, 2025.
[81] OpenAI. 2023. *DALL·E 3.* https://openai.com/index/dall-e-3/ Accessed: 2025-04-15.
[82] OpenAI. 2025. *ChatGPT.* https://chat.openai.com/ Accessed: 2025-04-15.
[83] Maulishree Pandey, Sharvari Bondre, Sile O'Modhrain, and Steve Oney. 2022. Accessibility of UI Frameworks and Libraries for Programmers with Visual Impairments. In *2022 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*. 1–10. doi:10.1109/VL/HCC53370.2022.9833098
[84] 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. doi:10.1145/3373625.3417082
[85] William Christopher Payne, Alex Yixuan Xu, Fabiha Ahmed, Lisa Ye, and Amy Hurst. 2020. How Blind and Visually Impaired Composers, Producers, and Songwriters Leverage and Adapt Music Technology. 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 35, 12 pages. doi:10.1145/3373625.3417002
[86] Yi-Hao Peng, Peggy Chi, Anjuli Kannan, Meredith Ringel Morris, and Irfan Essa. 2023. Slide Gestalt: Automatic Structure Extraction in Slide Decks for Non-Visual Access. 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 829, 14 pages. doi:10.1145/3544548.3580921

[87] Yi-Hao Peng, JiWoong Jang, Jeffrey P Bigham, and Amy Pavel. 2021. Say It All: Feedback for Improving Non-Visual Presentation Accessibility. 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 276, 12 pages. doi:10.1145/3411764.3445572

[88] 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. doi:10.1145/3526113.3545637

[89] Rusty Perez. 2024. *Big flaw with Meta AI and the glasses.* https://www.applevis.com/forum/apple-hardware-compatible-accessories/big-flaw-meta-ai-glasses Accessed: 2025-04-15.

[90] Venkatesh Potluri, Tadashi E Grindeland, Jon E. Froehlich, and Jennifer Mankoff. 2021. Examining Visual Semantic Understanding in Blind and Low-Vision Technology Users. 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 35, 14 pages. doi:10.1145/3411764.3445040

[91] Venkatesh Potluri, Liang He, Christine Chen, Jon E. Froehlich, and Jennifer Mankoff. 2019. A Multi-Modal Approach for Blind and Visually Impaired Developers to Edit Webpage Designs. In *Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility* (Pittsburgh, PA, USA) *(ASSETS '19)*. Association for Computing Machinery, New York, NY, USA, 612–614. doi:10.1145/3308561.3354626

[92] Chenyang Qi, Xiaodong Cun, Yong Zhang, Chenyang Lei, Xintao Wang, Ying Shan, and Qifeng Chen. 2023. Fatezero: Fusing attentions for zero-shot text-based video editing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 15932–15942.

[93] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning*. PmLR, 8748–8763.

[94] Google Research. 2023. *VideoPoet: A Large Language Model for Zero-Shot Video Generation.* https://research.google/blog/videopoet-a-large-language-model-for-zero-shot-video-generation/ Accessed: 2025-04-15.

[95] Nina Reviers and Sabien Hanoulle. 2023. Aesthetics and participation in accessible art experiences: Reflections on an action research project of an audio guide. *Journal of Audiovisual Translation* 6, 2 (2023), 99–121.

[96] Tone Roald and Johannes Lang. 2013. *Art and identity: Essays on the aesthetic creation of mind*. Vol. 32. Rodopi.

[97] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 10684–10695.

[98] Ethan Z. Rong, Mo Morgana Zhou, Zhicong Lu, and Mingming Fan. 2022. "It Feels Like Being Locked in A Cage": Understanding Blind or Low Vision Streamers' Perceptions of Content Curation Algorithms. In *Proceedings of the 2022 ACM Designing Interactive Systems Conference* (Virtual Event, Australia) *(DIS '22)*. Association for Computing Machinery, New York, NY, USA, 571–585. doi:10.1145/3532106.3533514

[99] Abir Saha and Anne Marie Piper. 2020. Understanding Audio Production Practices of People with Vision Impairments. In *Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility* (Virtual Event, Greece) *(ASSETS '20)*. Association for Computing Machinery, New York, NY, USA, Article 36, 13 pages. doi:10.1145/3373625.3416993

[100] Richard Sandell, Jocelyn Dodd, and Rosemarie Garland-Thomson. 2010. Representing disability. *Activism and agency in the museum. London* (2010).

[101] 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. doi:10.1145/3411764.3445242

[102] Woosuk Seo and Hyunggu Jung. 2017. Exploring the Community of Blind or Visually Impaired People on YouTube. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility* (Baltimore, Maryland, USA) *(ASSETS '17)*. Association for Computing Machinery, New York, NY, USA, 371–372. doi:10.1145/3132525.3134801

[103] Woosuk Seo and Hyunggu Jung. 2021. Understanding the community of blind or visually impaired vloggers on YouTube. *Univers. Access Inf. Soc.* 20, 1 (March 2021), 31–44. doi:10.1007/s10209-019-00706-6

[104] Shawn Shan, Jenna Cryan, Emily Wenger, Haitao Zheng, Rana Hanocka, and Ben Y. Zhao. 2023. Glaze: Protecting Artists from Style Mimicry by Text-to-Image Models. In *32nd USENIX Security Symposium (USENIX Security 23)*. USENIX Association, Anaheim, CA, 2187–2204. https://www.usenix.org/conference/usenixsecurity23/presentation/shan

[105] Shawn Shan, Wenxin Ding, Josephine Passananti, Stanley Wu, Haitao Zheng, and Ben Y. Zhao. 2024. Nightshade: Prompt-Specific Poisoning Attacks on Text-to-Image Generative Models. In *2024 IEEE Symposium on Security and Privacy (SP)*. 807–825. doi:10.1109/SP54263.2024.00207

[106] Ather Sharif, Venkatesh Potluri, Jazz Rui Xia Ang, Jacob O. Wobbrock, and Jennifer Mankoff. 2024. Touchpad Mapper: Examining Information Consumption From 2D Digital Content Using Touchpads by Screen-Reader Users. In *Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility* (St. John's, NL, Canada) *(ASSETS '24)*. Association for Computing Machinery, New York, NY, USA, Article 128, 4 pages. doi:10.1145/3663548.3688505

[107] Shelly Sheynin, Adam Polyak, Uriel Singer, Yuval Kirstain, Amit Zohar, Oron Ashual, Devi Parikh, and Yaniv Taigman. 2024. Emu edit: Precise image editing via recognition and generation tasks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 8871–8879.

[108] Abigale Stangl, Ann Cunningham, Lou Ann Blake, and Tom Yeh. 2019. Defining Problems of Practices to Advance Inclusive Tactile Media Consumption and Production. In *Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility* (Pittsburgh, PA, USA) *(ASSETS '19)*. Association for Computing Machinery, New York, NY, USA, 329–341. doi:10.1145/3308561.3353778

[109] Abigale Stangl, Jeeeun Kim, and Tom Yeh. 2014. 3D printed tactile picture books for children with visual impairments: a design probe. In *Proceedings of the 2014 Conference on Interaction Design and Children* (Aarhus, Denmark) *(IDC '14)*. Association for Computing Machinery, New York, NY, USA, 321–324. doi:10.1145/2593968.2610482

[110] 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. doi:10.1145/3441852.3471233

[111] Jennifer Sullivan Sulewski, Heike Boeltzig, and Rooshey Hasnain. 2012. Art and disability: Intersecting identities among young artists with disabilities. *Disability Studies Quarterly* 32, 1 (2012).

[112] Yujie Sun, Dongfang Sheng, Zihan Zhou, and Yifei Wu. 2024. AI hallucination: towards a comprehensive classification of distorted information in artificial intelligence-generated content. *Humanities and Social Sciences Communications* 11, 1278 (27 9 2024). doi:10.1057/s41599-024-03811-x Accessed: 2025-04-15.

[113] Pablo PL Tinio. 2013. From artistic creation to aesthetic reception: The mirror model of art. *Psychology of Aesthetics, Creativity, and the Arts* 7, 3 (2013), 265.

[114] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. 2023. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971* (2023).

[115] Dani Valevski, Matan Kalman, Eyal Molad, Eyal Segalis, Yossi Matias, and Yaniv Leviathan. 2023. Unitune: Text-driven image editing by fine tuning a diffusion model on a single image. *ACM Transactions on Graphics (TOG)* 42, 4 (2023), 1–10.

[116] Tess Van Daele, Akhil Iyer, Yuning Zhang, Jalyn C Derry, Mina Huh, and Amy Pavel. 2024. Making Short-Form Videos Accessible with Hierarchical Video Summaries. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 895, 17 pages. doi:10.1145/3613904.3642839

[117] 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. doi:10.1145/2818048.2820013

[118] Teng Wang, Jinrui Zhang, Junjie Fei, Hao Zheng, Yunlong Tang, Zhe Li, Mingqi Gao, and Shanshan Zhao. 2023. Caption Anything: Interactive Image Description with Diverse Multimodal Controls. arXiv:2305.02677 [cs.CV] https://arxiv.org/abs/2305.02677

[119] Wenxuan Wang, Haonan Bai, Jen-tse Huang, Yuxuan Wan, Youliang Yuan, Haoyi Qiu, Nanyun Peng, and Michael Lyu. 2024. New Job, New Gender? Measuring the Social Bias in Image Generation Models. In *Proceedings of the 32nd ACM International Conference on Multimedia* (Melbourne VIC, Australia) *(MM '24)*. Association for Computing Machinery, New York, NY, USA, 3781–3789. doi:10.1145/3664647.3681433

[120] Yukun Wang, Longguang Wang, Zhiyuan Ma, Qibin Hu, Kai Xu, and Yulan Guo. 2024. VideoDirector: Precise Video Editing via Text-to-Video Models. *arXiv preprint arXiv:2411.17592* (2024).

[121] Nicholas Wolterstorff. 1980. *Art in Action: Twards a Christian Aesthetic*. Wm. B. Eerdmans Publishing.

[122] 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* (Portland, Oregon, USA) *(CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 1180–1192. doi:10.1145/2998181.2998364

[123] Lotus Zhang, Jingyao Shao, Augustina Ao Liu, Lucy Jiang, Abigale Stangl, Adam Fourney, Meredith Ringel Morris, and Leah Findlater. 2022. Exploring Interactive Sound Design for Auditory Websites. 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 222, 16 pages. doi:10.1145/3491102.3517695

[124] Lotus Zhang, Abigale Stangl, Tanusree Sharma, Yu-Yun Tseng, Inan Xu, Danna Gurari, Yang Wang, and Leah Findlater. 2024. Designing Accessible Obfuscation Support for Blind Individuals' Visual Privacy Management. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 235, 19 pages. doi:10.1145/3613904.3642713

[125] 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* (New York, NY, USA) *(ASSETS '23)*. Association for Computing Machinery, New York, NY, USA, Article 8, 15 pages. doi:10.1145/3597638.3608387

[126] Mingrui Ray Zhang, Mingyuan Zhong, and Jacob O. Wobbrock. 2022. Ga11y: An Automated GIF Annotation System for Visually Impaired Users. 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 197, 16 pages. doi:10.1145/3491102.3502092

[127] Zhuohao (Jerry) Zhang and Jacob O. Wobbrock. 2023. A11yBoard: Making Digital Artboards Accessible to Blind and Low-Vision Users. 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 55, 17 pages. doi:10.1145/3544548.3580655

[128] Kaixing Zhao, Rui Lai, Bin Guo, Le Liu, Liang He, and Yuhang Zhao. 2024. AI-Vision: A Three-Layer Accessible Image Exploration System for People with Visual Impairments in China. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 8, 3, Article 145 (Sept. 2024), 27 pages. doi:10.1145/3678537

[129] 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. doi:10.1145/3134756

[130] Yu Zhong, Walter S. Lasecki, Erin Brady, and Jeffrey P. Bigham. 2015. RegionSpeak: Quick Comprehensive Spatial Descriptions of Complex Images for Blind Users. 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, 2353–2362. doi:10.1145/2702123.2702437