Audio Description Automatons: Exploring Perspectives on Personas for Generative AI Description Writing Assistants

Lucy Jiang
University of Washington
Seattle, WA, USA
lucjia@uw.edu

Amanda Zhu
UC San Diego
La Jolla, CA, USA
a5zhu@ucsd.edu

Brett Oppegaard
University of Hawaii at Manoa
Honolulu, HI, USA
brett.oppegaard@hawaii.edu

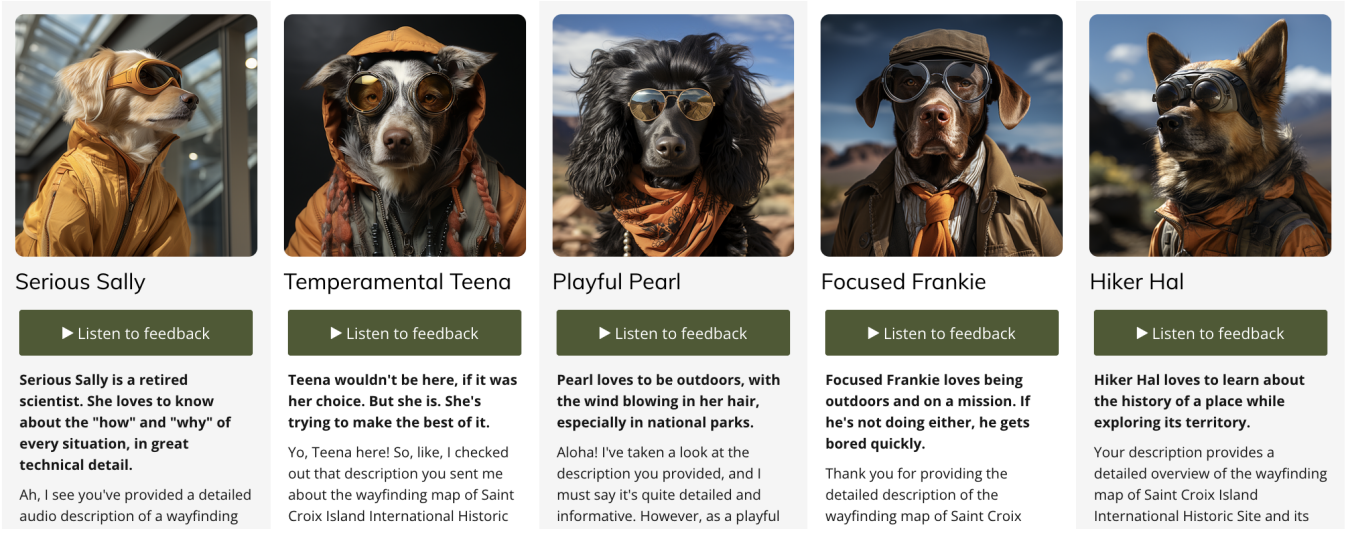

Figure 1: The five Guidedogs: Serious Sally, Temperamental Teena, Playful Pearl, Focused Frankie, and Hiker Hal. Each Guidedog has a unique photo and a short blurb explaining their personality.

Abstract

Visual media is often made accessible to blind and low vision (BLV) people through audio description (AD), typically written by experts. Prior efforts to increase the scale of description output have involved sighted novices as describers or used generative AI (GenAI) to automatically convert images to text; however, description quality remains a concern. To support novice describers in writing high quality descriptions, we designed and developed a GenAI-powered online tool, "Guidedogs," featuring five dogs with unique names, images, and voices that provided immediate and varied feedback on draft descriptions. We piloted the tool during a large hackathon-style description workshop in 2024. Through 17 semi-structured interviews, we explored the efficacy of using metaphors as personas for AI assistants and gathered insights on participants' perceptions on using AI for accessibility purposes. We contribute preliminary insights on generative AI assistant personas in an accessibility context and share design considerations to guide future work.

CHI EA '25, Yokohama, Japan

ACM ISBN 979-8-4007-1395-8/25/04
https://doi.org/10.1145/3706599.3719966

CCS Concepts

• **Human-centered computing** → **Accessibility**.

Keywords

audio description, image description, generative AI, personas, blind, low vision, AI assistant

ACM Reference Format:
Lucy Jiang, Amanda Zhu, and Brett Oppegaard. 2025. Audio Description Automatons: Exploring Perspectives on Personas for Generative AI Description Writing Assistants. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3706599.3719966

1 Introduction

Visual media is proliferating at an ever-increasing pace. For example, more than 400 National Park Service (NPS) sites across the United States use printed brochures with images to welcome, orient, and inform visitors. However, at present, most visual information is inaccessible to people who are blind, low vision, or DeafBlind.[1]

[1]We use person-first and identity-first language interchangeably to acknowledge and respect varied language preferences within the community [9, 30].

Images are typically made accessible through audio descriptions,[2] which can include both general and detailed information about the subject and other relevant image content [3, 32, 33, 35].

Audio descriptions (AD) can be written by trained professionals to ensure quality, but there are operational and economic barriers to having a small set of experts describe the large amount of visual content in our world. As such, some researchers have worked to engage sighted novices in writing AD, as involving the general public can increase the scale of description output (e.g., [24, 26, 29]). However, since they do not have the same training as professionals, sighted novices require additional guidance and / or feedback to improve their description quality. Other researchers have experimented with generative AI (GenAI) approaches for creating AD. For example, prior work has used AI to prompt novice describers to include specific details in their descriptions [26] or to enable BLV people to determine what types of descriptions they receive [8]. However, most have positioned GenAI systems as a replacement for human describers, rather than as a support system or assistant.

There remains a gap in understanding how GenAI-powered tools can support describers in improving description quality. Thus, we investigated the following research question: **How can we design personas for generative AI assistants to provide high-quality and effective feedback on audio description drafts?**

To assess the potential of AI-powered tools for novice audio description efforts, we designed and developed "Guidedogs," a set of five GenAI assistants that provided automated feedback on user-submitted descriptions. Unlike most GenAI interfaces that prompt back-and-forth conversations, we prompted the system to function as a writing coach and provide unidirectional feedback regarding best practices for audio descriptions. We also took a hypermediate approach to building the identities of each Guidedog persona, aligning with prior work on using personas to predict how audiences of different backgrounds will respond [13, 15].

The Guidedogs tool was piloted during a large hackathon-style description workshop (Descriptathon 10 [34]) focused on creating descriptions for U.S. National Park Service brochures. After the workshop concluded, we conducted semi-structured interviews with 17 participants to understand their motivations for using (or not using) the tool. Those who used the tool appreciated the design of the AI assistant and found the guide dog metaphor to be a fun and creative way to humanize the system. However, tool usage was not universal — across the interviews, participants had mixed perceptions about the usage of GenAI for accessibility purposes.

In this late-breaking work, we focus on the design rationale and development of the Guidedogs tool and share preliminary insights and reactions from workshop participants. Through the context of using AI to help non-experts learn about and support accessibility efforts, we aim to inform the development of AI-supported writing feedback systems more broadly. Our contributions include: (1) the design of a GenAI system using personas to provide feedback on novice descriptions, (2) findings from semi-structured interviews with 17 participants about the description workshop experience, and (3) design recommendations for future generative AI writing assistance systems.

[2]In this work, in alignment with terminology used by the National Park Service [27], we use the terms "audio description" (AD) and "description" to refer to static audio descriptions (also known as alt text or image descriptions).

2 Related Work

Our work builds on prior work regarding audio descriptions, automated and AI-supported tools for accessibility, and involving novices in creating access.

As AI technology has advanced in recent years, researchers have investigated how to harness its power for increasing AD output. For example, in an early exploration of using AI to create descriptions at scale, Wu et al. [37] evaluated user satisfaction with automatically generated descriptions for images on Facebook. They found that most participants were enthusiastic about having more context while viewing photos on social media, but some were still dissatisfied due to the vagueness and inaccuracies in the AI-generated descriptions. More recently, researchers have assessed the efficacy of GenAI for automatically creating detailed descriptions (e.g., [4, 6–8, 16, 25, 36]). However, despite major advancements in AD quality, BLV people expressed their desire for more humanized and less robotic descriptions [7].

Researchers have also explored how to scaffold and support non-experts in writing effective descriptions [20, 23, 38]. For example, Morash et al. [23] assessed novice performance for free response descriptions compared to queried descriptions (created by querying the describer for key information, then inputting the information into a template). Through a study with 22 novice describers, they identified that queried descriptions were more detailed and standardized than their free-response counterparts, suggesting that guidance improved novices' description outcomes. Mack et al. [20] found that novice describers appreciated having some degree of support when writing descriptions. Additionally, they found that participants wrote lower quality descriptions when they started from automatic alt text compared to starting from scratch, highlighting potential harms associated with integrating AI in AD scaffolding efforts. With this concern in mind, we approached this intervention by leveraging GenAI as a way to prompt further thought and revisions, rather than prescribe finished products.

A few additional studies have explored the intersection of automated feedback processes to support novice describers. For example, to aid novices in writing comprehensive and detailed video descriptions, Natalie et al. [26] used video scene recognition and natural language processing techniques to identify additional subjects that a novice could include in their description. They found that their automated feedback system could improve the descriptiveness and objectivity of the resulting descriptions, demonstrating the viability of AI-assisted description authoring tools for novices.

To our knowledge, no research has investigated how GenAI-powered systems can use multiple personas to simultaneously provide diverse feedback on both the content and style of draft audio descriptions. In particular, we aim to actively include novices as stakeholders and use AI as a support rather than a replacement for human involvement.

3 Guidedogs System

Our Guidedogs tool was designed to be used during the Descriptathon workshops and to integrate with existing open-access and open-source software created by our team. Our objective was to design a GenAI tool that guided novices and provided useful feedback in an engaging and personable way. We developed five personas for

the Guidedogs system: Serious Sally, Temperamental Teena, Playful Pearl, Focused Frankie, and Hiker Hal. Our design choices centered on two primary aspects: the form (as the metaphor of guide dogs, not as a park ranger or other human-like assistant) and the function (as audience feedback, not as a ghost writer).

3.1 Form Design

We chose to use **guide dogs** as a metaphor and aesthetic concept for our AI system for two primary reasons. Firstly, guide dogs are common companions and mobility aids for BLV people [31]. Secondly, we opted for dogs rather than human-like personas to highlight the hypermediacy of our system and clarify that the advice dispensed by the tool was computational and unmonitored. To make the personas more memorable, we assigned them names, photo-realistic images of dogs wearing clothing (generated via text-to-image platform Midjourney [22]), and unique voices for screen reader users (generated via text-to-speech platform Mimic [2]).

For the personas themselves, we designed five personas of archetypal NPS visitors to reflect the diversity of feedback that BLV visitors might offer. For example, Temperamental Teena had the perspective of a bored teenager, Focused Frankie valued conciseness, and Hiker Hal preferred knowing more about history. We chose to create multiple personas, as opposed to fine-tuning a single persona, to assess the efficacy of multiple feedback streams in an audio description context. While we acknowledge that this relatively small set of personas cannot comprehensively represent all BLV people's perspectives, we ultimately selected five Guidedogs to avoid information overload and feedback delays. Figure 1 shows the feedback interface and images of each Guidedog persona.

3.2 Function Design

As the development of the Guidedogs tool constituted the first time that the research team had integrated GenAI into the multi-day description workshop, we drew on insights from academic literature to design the system's functionality. For example, Gero et al. [11] identified that writers sought support for planning, translation, reviewing, and motivation. Writers also valued intention, authenticity, and creativity from those who were supporting them. Most commonly, writers wanted help reviewing the writing they already did; as a result, we designed the tool to coach novice writers rather than to replace them. This approach also aligns with BLV community guidance and prior work that advocates for maintaining human involvement in description efforts (e.g., [6, 20]).

We chose to design the tool to provide detailed and supportive critique to existing descriptions, but intentionally did not support dialectical discussion as GenAI systems may not be able to perform adequately in such a context [4]. The system used the GPT-3.5 Turbo API [28], which was the latest version available at the time of the workshop (February 2024). Each Guidedog persona was engineered through a prompt template developed by the research team. The assistant's personality was created through instructions that established a backstory as well as a communication style; for example, one prompt directed the AI assistant to open her response with phrasing that was "pleasant and colloquial, not overly formal." Our prompts also dictated the response's point of view (i.e., first-person) and the persona's general motivations and interests pertaining to description details. The assistants were all programmed to favor specific scholarly and official sources over anecdotal evidence on the internet. We also included prompts regarding the structure of the assistant's response in terms of content order, what types of advice should be given, and what types of advice should not. Each prompt added to the unique nature of the assistant and gave it a personality distinct from both other personas in the Guidedogs tool and external AI assistants (e.g., ChatGPT, Gemini, Copilot).

4 User Study

After the description workshop, we conducted a semi-structured interview study to assess the efficacy of our system and investigate our overarching research question: How can we design personas for generative AI assistants to provide high-quality and effective feedback on audio description drafts? Through our interviews, we identified two emerging sub-research questions:

- What are describers' perspectives on the use of stylized GenAI personas as AD-writing support systems?
- What are describers' perspectives on GenAI in terms of its potential to be integrated into the AD process?

4.1 Background and Participants

As this tool was only used during our multi-day description workshop, we recruited directly from this group of approximately 160 participants. The workshop consisted of 16 teams of 10 people on average, and each team was assigned to describe one U.S. National Park Service brochure or materials for an aquarium or zoo. Most workshop participants were sighted volunteers based in the United States, but some were from Canada, Italy, Mexico, or the United Kingdom. Some workshop participants were also members of the BLV community.

We began recruitment by reaching out to all 16 team captains, as their role gave them a level of insight into all team members' experiences. Then, we used a snowball sampling method to recruit additional team members who actively used the Guidedogs or contributed to team discussions about them. We also specifically reached out to participants associated with the American Council of the Blind (ACB), the Canadian Council of the Blind (CCB), or the Royal National Institute of Blind People (RNIB) as we aimed to represent a global BLV community perspective. In total, 17 people (seven men and ten women) participated in our IRB-approved study. Of our participants, 10 were sighted team captains, two were sighted volunteers, and five were leaders from the BLV community who identified as blind, low vision, or DeafBlind.

4.2 Procedure

Our study involved a 20 - 45 minute semi-structured interview session, conducted via Zoom from March to June 2024. Participants were invited to discuss their experiences with and impressions of Guidedogs tool use during the workshop. Eight of the 17 participants directly used the tool during the event. The remaining participants either shared how their team members worked with the tool or described why they did not use the tool.

Participants who used the tool were prompted to share their reflections and impressions of the Guidedogs, including describing tool use within their team and how frequently the tool was used.

Participants with direct experience using the Guidedogs tool, about half of our pool, also responded to a series of Likert-type scale questions assessing the Guidedogs' quantity of feedback, level of focus, level of detail, and overall usefulness. Participants who had not used the tool were asked about their choice to not use the tool and why they made that decision.

We audio and video recorded all interviews. The interviews were transcribed by the research team, which consisted of three members. We took a grounded theory approach for data analysis. We first separated the texts of the transcripts into discrete thought units (1,673 total). We then completed a round of open coding, with two coders independently developing a code for each thought unit. Through discussion and deliberation, we developed a new codebook of seven axial codes that represented the themes underlying our open codes. With three coders analyzing all of the transcripts, we reached a 90% agreement rate among at least two of the three coders. Thought units with at least two-thirds agreement were then clustered into themes and analyzed to reach our final results.

4.3 Findings

Through our interviews and data analysis, we identified three primary clusters of comments about: (1) the persona design of our tool (e.g., the guide dog metaphor), (2) the application of GenAI to support writing for accessibility purposes, and (3) the usage and usability of the tool (Table 1). We also share statistics regarding usage of the Guidedogs tool as collected through our system's backend analytics as a way to triangulate our findings.

4.3.1 Thoughts on the Guide Dog Metaphor. Almost all participants shared that they liked the guide dog persona design of the AI tool due to their familiarity, relatability, and topical relevance. For example, P8 shared that they thought the general public could *"understand what the concept is, if you say guide dogs as opposed to any other animal."* Others also commented that the metaphor was *"fun and creative"* (P11).

Participants appreciated the usage of photos and names associated with the different Guidedogs, and found them to be helpful for forming an identity, humanizing the tool, and making them *"more likeable"* (P2). Some mentioned that these additional features helped them warm up to the AI tool over the course of the multi-day description event: *"in those three days, they had ceased to become robots and they had become dogs"* (P2). Others shared that they were initially skeptical of the Guidedogs tool, with *"these dogs dressed [up] and the names and everything"* (P12), but ultimately found them to be *"a cute way of making it fun and entertaining... it was good to have some character to it"* (P12).

4.3.2 Thoughts on Generative AI. Participants were divided on whether they were comfortable using AI tools for accessibility purposes. While some were enthusiastic about AI's ability to provide rapid feedback for iteration and learning, others felt wary about potential pitfalls with regards to ethics, quality, and accuracy.

Some participants mentioned that using the AI tool helped them describe more quickly and creatively, without compromising a description's overall quality. For example, P10 felt that AI was *"a force multiplier... assisting what is already being done."* Participants also shared that the multiple personas associated with the Guidedogs tool helped them simulate the experience of receiving feedback from BLV audiences with diverse preferences: *"it provided specific examples of different audiences that we are trying to reach... it led to discussion among the team"* (P17). P5, who used AI tools for writing support during her day job, used multiple external AI services during the workshop, such as ChatGPT, Gemini, and Hemingway. She described her usual AI workflow: *"generally I like to write my own thing, and then play with whatever things come up in AI and what those Guidedogs would say"* (P5). Despite having experience with multiple LLM-based writing tools, she emphasized that her goal was to use the tools intentionally and somewhat sparingly, as she viewed AI tools as useful for copy editing and fine tuning.

Other participants were less enthusiastic about using AI. For example, P2 shed doubt on its accuracy: *"I'm like, ugh AI... I don't trust you, robot."* P14, who used multiple LLM writing tools during the workshop, shared an anecdote where the AI system hallucinated and *"insisted that one of these lizards I was describing does not have a tail."* Although she caught the mistake, she mentioned the harms of overreliance on AI: *"sometimes if you don't point it out... it keeps building on this myth"* (P14). Participants also expressed concerns about plagiarism, abuse of the system, and AI's removal of humanity. As someone who was fairly new to AI and still figuring out her stance, P16 expressed that she was optimistic about AI's potential but also could not overcome her worry of AI systems *"taking over and the human component of things being lost."*

P3, an interview participant who was blind, had particularly mixed feelings about AI. In her daily life, she sometimes used AI-supported applications such as Envision [10] and Apple's Siri for personal and experimental purposes. However, she felt that it was inappropriate to use AI as part of the Descriptathon, a stance which was influenced by her role during the workshop (as a paid professional providing feedback on novice-written descriptions). Out of concerns regarding professionality and accuracy, she stated that she didn't *"feel comfortable, at this point, using AI technology [during the workshop]" and "wouldn't rely on something unless [she] knew that it was sophisticated enough"* (P3). Her criteria for adopting AI technologies for creating publishable descriptions were that they would need to *"have the same ethical, same equivalent moral compass that a human should have"* (P3) and be widely recognized as high-quality within the accessibility community.

4.3.3 Tool Usage and Usability. During the three-day workshop, per the backend data, the Guidedogs were used 59 separate times, generating a total of 295 unique responses from the five different personas. Interview participants who used the tool shared that they thought the Guidedogs were helpful for gathering feedback from diverse perspectives. Despite its benefits, some shared that having five Guidedog personas was overwhelming and provided too much feedback to reasonably address. In her attempts to cater her descriptions to all of the Guidedogs' preferences, P16 repeatedly entered her descriptions into the tool for feedback, but felt that the system was *"sending [her] in circles."* Others like P12 thought that having *"two different spins or interpretations of it helped when we got stuck"* but *"I don't know if we needed all five of them."* Overall, the Guidedogs provided diverse and helpful input, but they could occasionally cause participants to spend even more time over-optimizing or processing large volumes of feedback.

Table 1: Likert-type scale ratings shared by the eight participants who used the Guidedogs tool. The endpoints were: quantity of feedback (1 = not enough, 5 = too much); level of focus (1 = not focused, 5 = overly focused); level of detail (1 = not detailed, 5 = overly detailed); usefulness (1 = not useful, 5 = extremely useful). The nine participants who did not use the tool are P3 (BLV), P4 (BLV), P5 (S), P7 (S), P8 (S), P9 (BLV), P10 (S), P15 (BLV), and P17 (S), where (S) represents a participant who is sighted and (BLV) represents a participant who is blind, has low vision, or is DeafBlind.

P#	Level of Vision	Feedback	Focus	Detail	Usefulness
P1	Sighted	3	3	3	4
P2	Sighted	4	3	4	5
P6	BLV	3	4	4	5
P11	Sighted	4	3	3	4
P12	Sighted	3	3	3	4
P13	Sighted	3	4	4	4
P14	Sighted	3	3	3	4
P16	Sighted	5	3	3	4
Mean		3.5	3.25	3.375	4.25
St Dev		0.756	0.463	0.518	0.463
Median		3	3	3	4
Mode		3	3	3	4

5 Discussion

In this section, we present recommendations for the design of future AI tools assisting with description writing, based on our Guidedogs design and insights shared by our participants. We also briefly touch on the importance of engaging novices in access efforts. Lastly, we discuss limitations regarding the scope of our study and share how future work could focus on generalizing our findings to support description writing in additional domains.

5.1 Design Recommendations for AI Tools Assisting with Writing Descriptions

In our study, we found that both sighted and BLV participants had mixed opinions about AI technology. Some were enthusiastic early adopters, while others had serious concerns about plagiarism, ethics, and accuracy. Prior work has shown that some people who are blind or low vision have started to embrace AI tools in their daily lives (e.g., [1, 12]), often for increasing agency in accessing visual descriptions. However, when using AI to create official and published work (such as NPS brochures), there is still a pressing need for greater accuracy, verifiability, and trust prior to more widespread adoption of these AI tools.

Extending prior work on writing tools (e.g., [19]) and providing feedback for image describers (e.g., [20, 23, 38]), we provide three primary design recommendations for future AI tools for description-writing. Firstly, we suggest that future AI systems **integrate more ways for both sighted and BLV users to verify their output**. For example, this could take the form of clearly demarcating which parts of the image are referenced in the AI system's feedback. It is important to note that a user's trust of a system can be impacted by the quality of feedback as well as the interaction itself [21]. Secondly, we recommend that tools **offer multiple streams of feedback to simulate real-life scenarios**. Participants expressed that having multiple personas as part of one tool allowed them to parse through different perspectives of feedback (e.g., some Guidedogs preferred more details about nature, while others preferred more concise descriptions). They noted that this affordance was especially helpful, since it mirrored their real-world experiences of receiving diverse feedback from blind and low vision people. Lastly, we encourage AI systems to **leverage familiar metaphors and designs to support user adoption**, given participants' positive responses to the Guidedog design of our AI tool.

5.2 Engaging Novices in Access Efforts

Given the importance of visual content in both physical and digital spaces, it is critical to ensure that BLV people have access to this information; however, the ubiquity of this content also means that it is impractical to wait for professionals to describe it all. As such, we intentionally involved novices in our workshop to (1) teach them about the importance of accessibility and high-quality descriptions and (2) increase the scale of description output, similar to prior studies on crowdsourcing for image descriptions (e.g., [5, 14]).

Through our research, we have demonstrated the value of engaging sighted novices in access efforts such as our multi-day description workshop. Many novice describers recognize the importance of accessibility, but lack exposure to hands-on opportunities or trainings to improve their description-writing skills. Especially when trained and supported properly, novice involvement can help scale description efforts quickly and efficiently — over time, these novices may even become experts. We encourage researchers and practitioners to consider including sighted novices in access efforts to expand awareness and investment in accessibility. In line with prior work, we also strongly advocate for the active inclusion of blind and low vision people as description creators, stakeholders, and experts [8, 17, 18].

5.3 Limitations and Future Work

There are some limitations to our system. For example, the Guidedogs system only processed text input and output, meaning that it could not "view" the image being described. While participants did not encounter egregious output errors, as their description inputs provided sufficient content for critique, we acknowledge that accuracy could be improved with the integration of more advanced multimodal systems. Additionally, our system featured five personas predetermined by the research team, based on the context of describing National Park Service brochures. Future work could explore additional Guidedog personas to explore how GenAI systems could support description-writing across different contexts (e.g., live theater, STEM classes, sporting events, etc.).

We also have identified some limitations to our interview study. For example, we asked participants to answer Likert-type scale questions about the system during our interviews, which took place one to three months after the workshop. Additionally, as some BLV people participated in the workshop in a paid role to give feedback to novices (rather than as volunteer writers themselves), we did not interview as many BLV people with experience using the Guidedogs tool. As this is an emerging area, we encourage researchers to continue studying GenAI tool design and usage, as well as ways to support and encourage widespread involvement.

6 Conclusion

This work explores the initial design and reception of the Guidedogs tool, a generative AI system intended to provide feedback on static audio descriptions for images, maps, and collages. We first describe the design rationale behind our system, which builds on prior research on GenAI writing assistance, image and video descriptions, and involving novices in access efforts. Then, we present results from our user study with 17 participants, including five people who are blind, low vision, or DeafBlind. We found that participants appreciated the guide dog metaphor used to characterize the personas but had mixed opinions on generative AI usage. Especially as the AI landscape is constantly changing, we recommend for future work to further explore AI tool design for accessibility, with an emphasis on increasing both the quality and quantity of described content.

Acknowledgments

This work was supported by the U.S. National Endowment for the Humanities under Grant 276851-21; the U.S. National Park Service under Grant P20AC01084-01; and Google, under a grant administered via the UH Foundation, 127-7390-4. Lucy Jiang was supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-2139899.

References

[1] 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*. 1–14. https://doi.org/10.1145/3663548.3675631

[2] Mycroft AI. [n. d.]. *Mimic TTS*. https://mycroft-ai.gitbook.io/docs/mycroft-technologies/mimic-tts

[3] American Council of the Blind. 2021. *Describing race, skin color, ethnicity, gender and disability in today's audio description*. https://www.acbmedia.org/2021/08/24/20210722-describing-race-skin-color-ethnicity-gender-and-disability-in-todays-audio-description/

[4] Daniel Bergin and Brett Oppegaard. 2024. Automating Media Accessibility: An Approach for Analyzing Audio Description Across Generative Artificial Intelligence Algorithms. *Technical Communication Quarterly* (2024), 1–16. https://doi.org/10.1080/10572252.2024.2372771

[5] Jeffrey P Bigham, Chandrika Jayant, Hanjie Ji, Greg Little, Andrew Miller, Robert C Miller, Robin Miller, Aubrey Tatarowicz, Brandyn White, Samuel 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*. 333–342. https://doi.org/10.1145/1866029.1866080

[6] Aditya Bodi, Pooyan Fazli, Shasta Ihorn, Yue-Ting Siu, Andrew T Scott, Lothar Narins, Yash Kant, Abhishek Das, and Ilmi Yoon. 2021. Automated Video Description for Blind and Low Vision Users. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–7. https://doi.org/10.1145/3411763.3451810

[7] Ruei-Che Chang, Yuxuan Liu, and Anhong Guo. 2024. WorldScribe: Towards Context-Aware Live Visual Descriptions. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–18. https://doi.org/10.1145/3654777.3676375

[8] Maryam Cheema, Hasti Seifi, and Pooyan Fazli. 2024. Describe Now: User-Driven Audio Description for Blind and Low Vision Individuals. *arXiv preprint arXiv:2411.11835* (2024). https://arxiv.org/abs/2411.11835

[9] Dana S Dunn and Erin E Andrews. 2015. Person-first and identity-first language: Developing psychologists' cultural competence using disability language. *American Psychologist* 70, 3 (2015), 255. https://doi.org/10.1037/a0038636

[10] Envision. [n. d.]. *Envision - Perceive Possibility*. https://www.letsenvision.com/

[11] Katy Ilonka Gero, Tao Long, and Lydia B Chilton. 2023. Social dynamics of AI support in creative writing. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–15. https://doi.org/10.1145/3544548.3580782

[12] Ricardo E Gonzalez Penuela, Jazmin Collins, Cynthia Bennett, and Shiri Azenkot. 2024. Investigating Use Cases of AI-Powered Scene Description Applications for Blind and Low Vision People. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–21. https://doi.org/10.1145/3613904.3642211

[13] Kathleen W Guan, Joni Salminen, Soon-Gyo Jung, and Bernard J Jansen. 2024. Leveraging personas for social impact: A review of their applications to social good in design. *International Journal of Human–Computer Interaction* 40, 19 (2024), 5569–5584. https://doi.org/10.1080/10447318.2023.2247568

[14] Danna Gurari, Qing Li, Abigale J Stangl, Anhong Guo, Chi Lin, Kristen Grauman, Jiebo Luo, and Jeffrey P Bigham. 2018. Vizwiz grand challenge: Answering visual questions from blind people. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3608–3617. https://doi.org/10.1109/CVPR.2018.00380

[15] Alexander Henka and Gottfried Zimmermann. 2014. Persona Based Accessibility Testing: Towards User-Centered Accessibility Evaluation. In *HCI International 2014-Posters' Extended Abstracts: International Conference, HCI International 2014, Heraklion, Crete, Greece, June 22-27, 2014. Proceedings, Part II 16*. Springer, 226–231. https://doi.org/10.1007/978-3-319-07854-0_40

[16] 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 CHI Conference on Human Factors in Computing Systems*. 1–20. https://doi.org/10.1145/3613904.3642238

[17] Lucy Jiang and Richard Ladner. 2022. Co-Designing Systems to Support Blind and Low Vision Audio Description Writers. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–3. https://doi.org/10.1145/3517428.3550394

[18] Lucy Jiang, Mahika Phutane, and Shiri Azenkot. 2023. Beyond Audio Description: Exploring 360° Video Accessibility with Blind and Low Vision Users Through Collaborative Creation. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–17. https://doi.org/10.1145/3597638.3608381

[19] Mina Lee, Katy Ilonka Gero, John Joon Young Chung, Simon Buckingham Shum, Vipul Raheja, Hua Shen, Subhashini Venugopalan, Thiemo Wambsganss, David Zhou, Emad A Alghamdi, et al. 2024. A Design Space for Intelligent and Interactive Writing Assistants. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–35. https://doi.org/10.1145/3613904.3642697

[20] Kelly Mack, Edward Cutrell, Bongshin Lee, and Meredith Ringel Morris. 2021. Designing tools for high-quality alt text authoring. In *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility*. 1–14. https://doi.org/10.1145/3441852.3471207

[21] Siddharth Mehrotra, Chadha Degachi, Oleksandra Vereschak, Catholijn M Jonker, and Myrthe L Tielman. 2024. A systematic review on fostering appropriate trust in Human-AI interaction: Trends, opportunities and challenges. *ACM Journal on Responsible Computing* 1, 4 (2024), 1–45.

[22] Midjourney. [n. d.]. *Midjourney*. https://www.midjourney.com/home

[23] Valerie S Morash, Yue-Ting Siu, Joshua A Miele, Lucia Hasty, and Steven Landau. 2015. Guiding novice web workers in making image descriptions using templates. *ACM Transactions on Accessible Computing (TACCESS)* 7, 4 (2015), 1–21. http://dx.doi.org/10.1145/2764916

[24] Sawako Nakajima and Kazutaka Mitobe. 2024. Professional and novice audio describers: quality assessments and audio interactions. *The Journal of Specialised Translation* 42 (2024), 64–83. https://doi.org/10.26034/cm.jostrans.2024.5980

[25] Rosiana Natalie, Ruei-Che Chang, Smitha Sheshadri, Anhong Guo, and Kotaro Hara. 2024. Audio description customization. In *Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–19.

[26] Rosiana Natalie, Joshua Tseng, Hernisa Kacorri, and Kotaro Hara. 2023. Supporting Novices Author Audio Descriptions via Automatic Feedback. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–18. https://doi.org/10.1145/3544548.3581023

[27] National Park Service. 2019. *Accesssibility and Audiovisual Media FAQ.* https://www.nps.gov/subjects/hfc/accessibility-and-audiovisual-media-faq.htm

[28] Open AI. [n. d.]. *GPT-3.5 Turbo.* https://platform.openai.com/docs/models#gpt-3-5-turbo

[29] Brett Oppegaard and Michael K Rabby. 2022. Gamifying good deeds: User experience, agency, and values in play during a descriptathon. *Technical Communication* 69, 4 (2022), 27–43. https://doi.org/10.55177/tc124312

[30] Ather Sharif, Aedan Liam McCall, and Kianna Roces Bolante. 2022. Should I Say "Disabled People" or "People with Disabilities"? Language Preferences of Disabled People Between Identity- and Person-First Language. In *Proceedings of the 24th international ACM SIGACCESS conference on computers and accessibility*. 1–18. https://doi.org/10.1145/3517428.3544813

[31] Arielle M Silverman, Carlie R Rhoads, Elizabeth Bolander, and Kelly Bleach. 2022. *The Role of Guide Dogs in 2022 and Beyond: Findings from a GDB-AFB Research Partnership.* https://afb.org/guidedogs-research

[32] Smithsonian Design Museum. 2024. *Cooper Hewitt Guidelines for Image Description.* https://www.cooperhewitt.org/cooper-hewitt-guidelines-for-image-description/

[33] Joel Snyder. 2005. Audio description: The visual made verbal. In *International Congress Series*, Vol. 1282. Elsevier, 935–939.

[34] UniDescription. [n. d.]. *Descriptathon 10.* https://unidescription.org/descriptathon/d10

[35] UniDescription. 2024. *UniDescription Academy.* https://unidescription.org/unid-academy

[36] Gustav Verhulsdonck, Jennifer Weible, Danielle Mollie Stambler, Tharon Howard, and Jason Tham. 2024. Incorporating Human Judgment in AI-Assisted Content Development: The HEAT Heuristic. *Technical Communication* 71, 3 (2024), 60–72. https://doi.org/10.55177/tc286621

[37] Shaomei Wu, Jeffrey Wieland, Omid Farivar, and Julie Schiller. 2017. Automatic alt-text: Computer-generated image descriptions for blind users on a social network service. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 1180–1192. https://doi.org/10.1145/2998181.2998364

[38] 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*. 1–16. https://doi.org/10.1145/3491102.3502092