

# “A Little Guidedog Voice on Their Shoulder”: Evaluating Generative AI Writing Assistants for Audio Description Creation

ANONYMOUS AUTHOR(S)

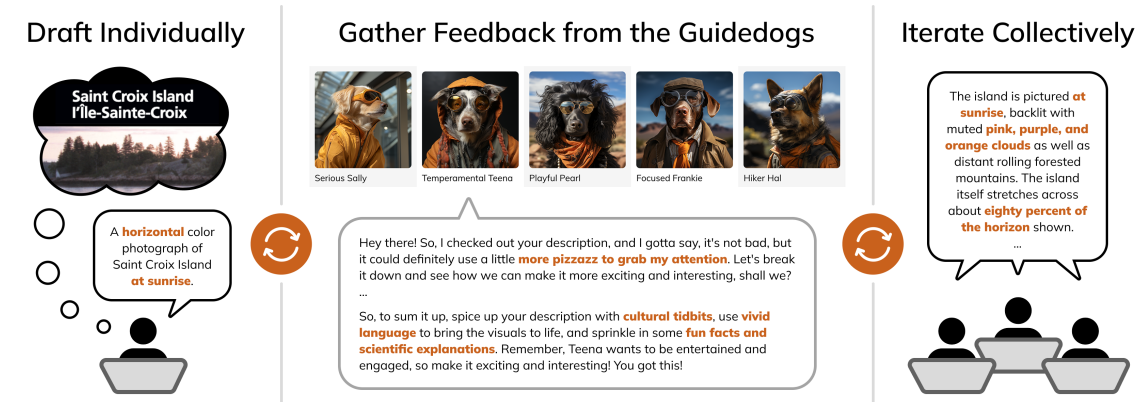


Fig. 1. An example workflow of using the Guidedogs tool during the description workshop begins with drafting descriptions individually, gathering feedback from the Guidedogs, then iterating collectively as a team. Participants often used the Guidedogs tool for multiple rounds of drafting and iteration.

Blind and low vision (BLV) people often use audio description (AD) to access information encoded in visual media, such as images or videos. While AD is typically created by trained writers, the pace of visual media production far exceeds current accessibility efforts. Researchers have explored how to involve novices in AD creation and how generative AI (GenAI) can automatically provide descriptions. However, we have yet to understand how GenAI can or should support novices to improve AD quality and quantity. We designed and developed “Guidedogs,” a GenAI-powered tool featuring five dogs with unique names, images, and voices that provided immediate and diverse feedback on draft AD. We piloted the tool during a large hackathon-style description workshop in 2024. We then conducted 17 semi-structured interviews with workshop participants to understand the usefulness, affordances, and constraints of the Guidedogs. While the tool provided unique affordances for novice AD creators, we surfaced emerging ethical concerns regarding GenAI and accessibility. We conclude by discussing design recommendations for future GenAI tools to support access efforts.

CCS Concepts: • Human-centered computing → Accessibility.

Additional Key Words and Phrases: audio description, image description, generative AI, personas, blind, low vision, AI assistant

## ACM Reference Format:

Anonymous Author(s). 2025. “A Little Guidedog Voice on Their Shoulder”: Evaluating Generative AI Writing Assistants for Audio Description Creation. 1, 1 (April 2025), 22 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.

Manuscript submitted to ACM

## 1 Introduction

“The National Park Service is committed to ensuring that visitors with disabilities have equal opportunity to benefit from our facilities, programs, services, and activities whether they are indoors or outdoors.”  
— U.S. National Park Service [46]

Images in both physical and digital sources are typically made accessible to blind and low vision (BLV) people through audio descriptions.<sup>1</sup> Despite the prevalence of visual content in informational, educational, and entertainment materials, including throughout the U.S. National Park Service’s 400+ sites, most images are inaccessible to people who are blind, low vision, or DeafBlind.<sup>2</sup>

While visual media is exponentially increasing in quantity, the number of professionals dedicated to making these materials accessible to BLV people is relatively small. To both scale and accelerate audio description (AD) creation, some researchers have turned their attention to generative AI (GenAI) interventions. For example, researchers have used advanced GenAI systems to generate live visual descriptions of the real world [16] and create descriptions of various lengths for short-form videos [57]. In tandem, multiple applications have harnessed the power of GenAI to provide access to blind and low vision users in practice (e.g., Be My AI [5], Seeing AI [39], Aira [2]). From these steps, it is clear that GenAI has immense *potential* to improve accessibility outcomes.

Other researchers have investigated how to integrate GenAI into human description workflows. Prior work has used AI to prompt novice describers to include specific details in their descriptions [44], and others have explored human-in-the-loop designs for AI-generated descriptions [62, 63]. However, a majority of recent approaches have framed AI as a way to fully substitute the costly and time-intensive process of human description creation. However, many BLV people and organizations advocate for human involvement throughout the AD creation process as a form of quality control [16, 54, 62].

Considering the tradeoffs of increasing quality and decreasing costs, the accessibility community — which includes researchers, practitioners, and BLV people — has yet to come to a consensus on how to best leverage GenAI technologies to create high-quality and humanized AD. It is important to consider the perspectives of both the audio describers who are being supported and the BLV users who will consume the AD outputs. To address this gap, we investigated the research question: **What are audio describers’ and BLV people’s perspectives on the usefulness, affordances, and constraints of generative AI tools for supporting audio description creation processes?**

We designed and developed a GenAI system, the “Guidedogs,” which was piloted during a large hackathon-style description workshop in February 2024. Our “Guidedogs” system consisted of five GenAI assistants, each with a specific persona, that provided automated feedback on users’ descriptions, akin to a writing coach. After the workshop concluded, we conducted semi-structured interviews with 17 participants to understand their experiences with using the tool as part of the workshop. We present participants’ thoughts on the usage and utility of the AI tool itself, key constraints to GenAI for accessibility, and insights on the social context in which this tool was deployed. Participants generally had positive experiences using the Guidedogs to spark inspiration, gather feedback from different perspectives, and improve novices’ sense of privacy when iterating on their descriptions. However, participants also sometimes felt that the Guidedogs caused confusion with conflicting suggestions, and they shared ethical concerns regarding AI usage for accessibility purposes.

<sup>1</sup>In this work, in alignment with terminology used by the U.S. National Park Service [45], we use the terms “audio description” (AD) and “description” to refer to descriptions of static media (also referred to as alt text or image description).

<sup>2</sup>We use person-first and identity-first language interchangeably to acknowledge and respect varied language preferences within the community [20, 52].

In this work, we describe the design rationale underscoring our Guidedogs tool and dive into describers’ perspectives on the utility and affordances of the tool as it was used during a three-day description workshop. Our contributions include: (1) the design and development of a GenAI system using personas to provide feedback on novice descriptions, (2) detailed insights from 17 semi-structured interviews about the usage, utility, and affordances of our Guidedogs webtool, and (3) design implications for future AI-supported description writing efforts. We advocate for a human-in-the-loop approach for description creation to achieve both quality and quantity.

## 2 Related Work

Our work builds on prior research regarding AI for accessibility, novice involvement in access efforts, and GenAI writing support.

### 2.1 AI for Visual Accessibility

Researchers investigated how to harness AI technologies for increasing AD output even prior to the mainstream release of GenAI systems in 2022. For example, in an early exploration of using AI to create descriptions at scale, Wu et al. [60] evaluated user satisfaction with automatically generated descriptions for images on Facebook. While some BLV people were enthusiastic about having more context while viewing photos on social media, many were dissatisfied due to the vagueness and inaccuracies in the AI-generated descriptions, a finding corroborated by other work [25, 55]. As AI-generated descriptions became more widely available through applications such as Seeing AI [39] and Envision AI [22], researchers further investigated BLV users’ reactions to automated scene descriptions. They found that a description’s level of detail was correlated with its perceived trust and usefulness [33].

More recently, researchers have assessed the efficacy of GenAI for automatically creating descriptions for both videos and people’s surroundings (e.g., [11, 16, 17, 28, 43, 57, 58]). For example, Van Daele et al. [57] created a short-form video description system that leveraged both vision language models and large language models to generate descriptions. The authors found that only two videos had three or more errors in the short description; while this level of accuracy still falls short of expectations typical for human-written descriptions, participants demonstrated a significant improvement in video comprehension over the baseline interface. To enhance BLV people’s access to their surroundings, Chang et al. [16] developed WorldScribe, a system that provides BLV people with contextual and customized descriptions on-demand. While WorldScribe was largely effective and accurate, showing marked improvement compared to early implementations of AI for accessibility (e.g., [60]), participants still expressed their desire for more humanized and less robotic descriptions [16]. In tandem, off-the-shelf applications that previously relied on human descriptions, such as BeMyEyes, have integrated GenAI to scale description capabilities [5].

Others have evaluated the ethics of using AI for descriptions, especially given its potential to amplify biases when identifying and describing human subjects [7, 8, 13, 59]. While people who are blind or have low vision appreciated the potential for AI to increase access to image descriptions, they also had significant ethical concerns. As a result, researchers encouraged future applications of automation to focus on supporting human description writing efforts rather than replacing them altogether [7].

In this work, we assess the utility and efficacy of a GenAI system specifically designed to support human audio describers. Specifically, we focus on the context of describing U.S. National Park Service site brochures, which are an official resource (as opposed to the more informal nature of social media) and include highly contextual images of people, artifacts, and environments.

## 2.2 Involving Novices in Accessibility Efforts

Early instances of research involving novices in accessibility efforts focused on crowdsourcing at a large scale (e.g., [9, 10, 14, 51, 56]). Through VizWiz, Bigham et al. [9] recruited multiple human workers to provide answers to BLV users' visual questions in nearly real-time. The authors found that involving human novices could greatly increase the speed at which BLV people received answers to questions about their surroundings. However, the questions were limited in scope and novice describers did not require much training to provide accurate and satisfactory responses. As image and video content continues to increase on the web, the need for high-quality descriptions grows as well.

As a result, researchers have explored how to scaffold and support non-experts in writing effective and detailed descriptions [37, 41, 64]. For example, Morash et al. [41] 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. [37] 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 assist novice describers. For example, Jiang et al. [29] designed a system to support blind and low vision people in creating AD. While BLV people are familiar with descriptions, they are often perceived as passive consumers. Despite BLV participants' lack of prior experience in creating AD, they clearly identified key qualities that they valued in descriptions when writing AD as part of this study. Additionally, to aid novices in writing comprehensive and detailed video descriptions, Natalie et al. [44] 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.

## 2.3 Generative AI Writing Support

Many subareas of the human-computer interaction field have researched how generative AI and large language models (LLMs) may help or hinder human writing efforts. In this section, we highlight a subset of recent papers focused on creative writing support, professional writing and ethics, and the usage of AI-generated personas for writing and ideation.

Researchers have studied how people perceive and use GenAI writing support systems [15, 23, 24, 61? ]. For example, through interviews with 20 creative writers, Gero et al. [24] identified that writers sought support for planning, translation, reviewing, and motivation. They found that writers also valued intention, authenticity, and creativity when receiving support during their writing process. Additionally, other researchers found that writers appreciated LLM support for translation (e.g., targeted rewriting of paragraphs) and reviewing (e.g., obtaining feedback), and expressed frustration at key limitations of systems such as their lack of nuance and overreliance on cliches [15]. Yeh et al. [61] designed GhostWriter, a GenAI-powered system to help users refine or generate text. Through interviews with 18 professionals in the technology industry, the authors found that participants viewed AI as a tool, collaborator, or

both. While those who viewed the AI as a tool acknowledged the user’s role as a prompter and instructor, those who perceived the AI as a collaborator emphasized its utility for idea generation and potential opportunities for back and forth engagement.

Others have focused more specifically on the ethics of using GenAI, especially in professional contexts. Otmar et al. [50] posited that ChatGPT and other LLMs could play a useful role in correcting minor errors and even supporting early brainstorming efforts, but also acknowledged key ethical issues such as copyright infringement and its threat to creators’ livelihoods. Furthermore, generative AI is known to hallucinate and fabricate information, which can have serious consequences [3]. To mitigate concerns about perpetuating falsehoods through AI-generated text, Laban et al. [34] proposed a framework to alert authors to problematic content introduced by LLMs. While this framework helped participants detect and avoid factual inaccuracies in their finished work, the samples used during this study were relatively short and could only verify publicly known information.

Some have also examined how personas can provide diverse feedback to writers. For example, Benharrak et al. [6] investigated writers’ perceptions of AI personas for receiving feedback from their target group. They found that the personas were helpful for gathering different perspectives, but writers struggled with developing and tailoring the personas themselves. Others developed PersonaFlow, which leveraged LLMs to provide feedback from the perspective of multiple personas during the ideation phase of research [36]. The authors identified a positive correlation between the number of personas used and the perceived creativity of the output, as the diversity of personas could provide valuable insights into new domains with which the author themselves was not familiar. While these works provide valuable foregrounding for understanding the role of personas in writing and ideation, most prior work on writing support and personas has focused on proficient or expert end users; less is known about the needs and experiences of novice creators working in an accessibility context.

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

The Guidedogs tool was designed to be used during the description 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 anthropomorphic dog personas for the Guidedogs system: Serious Sally, Temperamental Teena, Playful Pearl, Focused Frankie, and Hiker Hal. Figure 2 shows the feedback interface and images of each Guidedog, and some examples of their feedback are provided in Appendix A. 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. First, guide dogs are common companions and mobility aids for BLV people [53]. Second, we opted for dogs rather than human-like personas to highlight the hypermediacy of our system. By choosing non-human personas, we aimed to clarify that the advice dispensed by the tool was algorithmically generated and unmonitored. To make the personas more memorable, we assigned them names, photo-realistic images of anthropomorphic dogs wearing clothing (generated via text-to-image

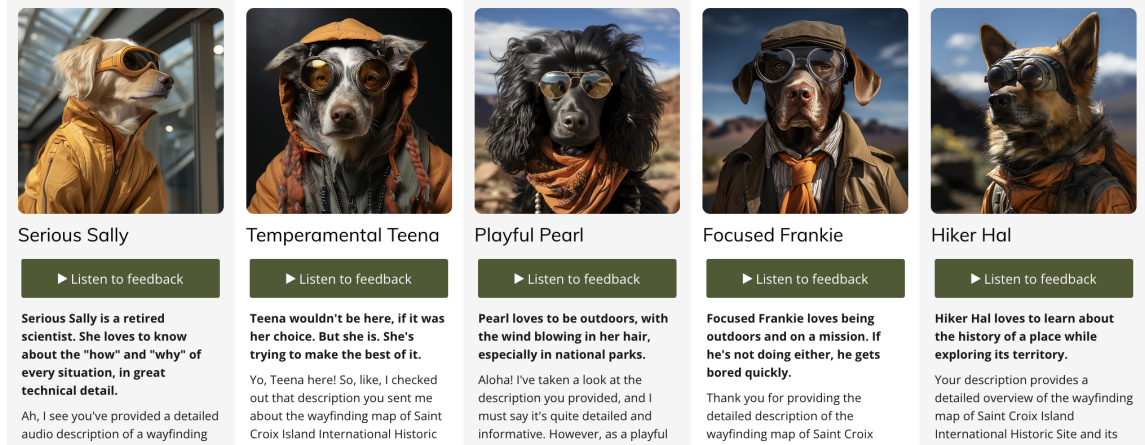


Fig. 2. The five Guidedogs: Serious Sally, Temperamental Teena, Playful Pearl, Focused Frankie, and Hiker Hal. Each Guidedog has a unique photo and a short blurb explaining their personality.

platform Midjourney [40]), and unique voices for screen reader users (generated via text-to-speech platform Mimic [42]).

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 for AD creation. 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.

### 3.2 Function Design

We drew on insights from academic literature (e.g., [24]) to design the Guidedogs tool's functionality. Given writers' values of authenticity and creativity, 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 AD efforts [11, 37].

We chose to design the tool to provide detailed and supportive critique to existing descriptions, but intentionally did not support dialectical discussion on AD as prior work identified that GenAI systems at the time could not necessarily perform adequately in such a context [8]. We used the GPT-3.5 Turbo API [48], which was the latest version available at the time of the workshop (February 2024). In line with the capabilities of the LLM, the Guidedogs system was only given the user's text description as input, meaning that it did not process or compare the current description to the image being described or any other previous descriptions. The tool generated responses based solely on a single description's content and wording.

Each Guidedog persona was engineered through a unique prompt template developed by the research team. We created each assistant's personality through instructions that established a backstory and 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 Methods

### 4.1 Background about the Description Workshop

The Guidedogs tool was first deployed as part of the 2024 description workshop, part of an ongoing series of hackathon-style audio description workshops that have been hosted approximately every year since 2016. The workshops typically require a 20-hour time commitment from volunteers, which include roughly 10 hours of asynchronous preparation and learning and a synchronous three-day period with an average of four hours of meetings per day. There is no cost for volunteers to participate in the description workshop and they receive access to a variety of free resources to learn more about AD (e.g., research papers, web-based training modules, etc.).

The description workshop features two portions: a bracket-style tournament encouraging describers to work together in teams to produce high-quality descriptions in a short amount of time, and a team-level task of describing a complete brochure for a U.S. National Park Service site. While the former is oriented towards providing a friendly competition environment that motivates and engages novice volunteers to practice their AD skills, the latter ensures that the ultimate aim of making NPS sites more accessible is met as well.

The 2024 workshop consisted of 16 teams of 10 people on average. In general, each team was assigned to describe one U.S. NPS 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. Most workshop participants were sighted volunteers, and audio description creation experience varied greatly — a majority did not have prior exposure to AD, but a few were professional AD writers for public services (e.g., describer for a city museum, interpreter for the NPS). Some workshop participants were also members of the BLV community, often serving in a paid role as a Judge or Justice to provide detailed feedback on the descriptions submitted as part of the tournament component of the workshop. Each team also had at least one Judge or Justice to provide feedback during the brochure description-writing process.

The Guidedogs tool was made available to all workshop participants, but usage was optional, allowing teams and individuals to decide whether and how to incorporate AI-generated feedback into their processes.

### 4.2 Participants

As this tool was only used during our multi-day description workshop, we recruited interview participants directly from this group of approximately 160 participants.

We began interview recruitment by reaching out to all 16 team captains from the description workshop, 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 workshop 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. Ultimately, all participants were involved in the 2024 description workshop; 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.

### 4.3 Procedure

The interview portion of 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 February 2024 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 (Table 1). We asked questions such as:

- What were your initial impressions of the Guidedogs tool?
- How often did you use the Guidedogs (every single time you wrote a description, sometimes, rarely, never)?
- What informed your decision to use the Guidedogs in certain situations?
- If you can recall, were there any times when you or your team directly used the feedback from the Guidedogs?

Participants who did not use the tool were asked about their choice to not use the tool and why they made that decision. If they did not use them because they were unaware of the tool, we described the Guidedogs feature and briefly presented each of the five personas. We then asked questions such as:

- If you can recall, do you believe a feature like this would have been helpful during any part of this last workshop?
- In a future workshop, would you use a similar feature? Why or why not?
- What suggestions would you have for how to improve the Guidedogs feature for a future workshop?

### 4.4 Data Analysis

We audio and video recorded all interviews, and the interviews were transcribed by the research team. We took a grounded theory approach for data analysis. We first divided the transcripts into discrete thought units (1,673 total), 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. Following the coding process, two authors conducted a reflexive thematic analysis on thought units with at least two-thirds agreement, based on guidance from Braun and Clarke [12, 18].

### 4.5 Positionality

All members of the research team are sighted. Multiple authors have experience collaborating with people who are blind or who have low vision. One of the authors has hosted over ten hackathon-style AD workshops since 2016, and another author has experience as an amateur AD creator for videos requested by friends in the BLV community. Notably, the



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
<b>Mean</b>		3.5	3.25	3.375	4.25
<b>St Dev</b>		0.756	0.463	0.518	0.463
<b>Median</b>		3	3	3	4
<b>Mode</b>		3	3	3	4

first author volunteered in the 2024 description workshop and had positive experiences with using the Guidedogs tool but did not serve as an interview participant.

## 5 Findings

From our interviews with description workshop participants, we identified that those who used the Guidedogs (N = 8) generally found them helpful for a variety of tasks including seeking writing inspiration and gathering feedback from different perspectives. However, some felt overwhelmed by the volume of guidance or confused by conflicting suggestions.

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. Tool usage was strongly influenced by the environment of the description workshop, as some participants mentioned that they wished to use the tool but simply did not have the chance during the three-day period, and others felt it was helpful for seeking support independently before having group discussions. Participants also shared thoughts about the guide dog metaphor and persona design of the tool, as well as broader thoughts about the role of GenAI in accessibility contexts. Our findings illuminate both the direct impact and outcomes of AI support on novice description writers as well as social and contextual factors that can determine when and how such tools can be incorporated into collaborative accessibility work.

### 5.1 Usage and Utility for Independent Writing Support

Participants generally found the Guidedogs useful for independent writing tasks, such as creating initial drafts, gathering feedback on existing drafts, and doing final passes of copyediting. In the following sections, we present specific examples of Guidedogs tool usage to demonstrate novice use cases during the description workshop.

5.1.1 *Writing Prompts and Inspiration.* Participants sought help from the Guidedogs for writing prompts and inspiration, whether that was to spark better ideas for initial drafts of descriptions or to improve fully drafted works. For example, to help with writer's block, P16 mentioned: *"I was just kind of stuck in a rut and wasn't really sure... where to go next with the description. So I was like, 'Okay, I'm just gonna throw this at the Guidedogs.'"* Participants also highlighted the Guidedogs' ability to provide word-level feedback, improving writing style and inspiring more engaging AD later on in the writing process. In particular, for novice describers, they found that it could improve previously *"bland and monotonous"* (P16) draft descriptions and *"spark a little more creativity in the writing"* (P16). Others mentioned that the Guidedogs could help with wording choice: *"one of the Guidedogs suggested, 'the ground is covered with a confetti of leaves.' I thought, 'that was a cool phrase.' And I took that piece"* (P1). Similarly, P11 shared how the Guidedogs gave her *"good food for thought"* and provided a variety of actionable changes such as reminders to shorten long descriptions and suggestions to include spatial information in descriptions.

While low-level edits were helpful, some also noted that more abstract feedback from the Guidedogs could be equally thought provoking for considering different approaches altogether. P12, a team captain affiliated with the National Park Service, mentioned: *"everyone comes in with their own way of describing things... [but] it was the Guidedog that did quite a... different take. ... 'Why not try this? Why not describe this way?'"* Even when participants did not agree with the Guidedogs' suggestions, they still found them useful for validating their current direction. For example, P16 disagreed with one of the Guidedogs' preference for *"more flowery language... I didn't like that, just get to the point."* Others shared that the Guidedogs' feedback sometimes *"reaffirmed a way I didn't want to go"* (P1). Overall, P13 acknowledged that the Guidedogs were especially useful given participants' general level of experience with writing AD: *"the fact that it is focused on audio description is helpful because that's a different medium that not most of us are working in."*

5.1.2 *Feedback from Different Perspectives.* Participants shared that one of the most useful aspects of the Guidedogs was their ability to provide feedback from five different end user perspectives: Serious Sally, Temperamental Teena, Playful Pearl, Focused Frankie, and Hiker Hal. P15 noted, *"what those Guidedogs illustrate, a little over abundantly perhaps, is the differences in the ways people approach audio description."* For example, P1 described how his usage of the Guidedogs enhanced his understanding of AD: *"if there's one thing that the [description workshop] taught me... there are a lot of different ways [to describe an image]... what the Guidedogs brought to the fore was that you can stylistically have different approaches."* P14 also found the multitude of perspectives helpful, especially for novices, because for those who are unfamiliar with AD, *"it's hard to realize how many perspectives there are from the blind community."* Each Guidedog had their own personality and perspective, but there were marked differences between each dog: *"the point of view was always consistent from the individual Guidedog, but there was no consensus of consistency among all Guidedogs"* (P1). For novice describers, the diversity of the Guidedog feedback encouraged them to consider different description styles.

Even when Guidedogs' feedback conflicted, participants described how this could still be helpful for making overall improvements. Participants often found themselves balancing different pieces of feedback from different Guidedogs. For example, P11 recalled that one of the Guidedogs provided feedback to include *"more analogies, or... really emotive language, and then a different Guidedog was like, 'This might be a little long already.' So I was like, 'Okay, there's some tension in that.' I feel like we were able to find a good compromise."* P14 noted how overlaps in otherwise divergent Guidedogs' feedback could also point describers towards best practices:

"Some personalities are focusing on 'get more storytelling' and others are more about '[going] into the cultural highlight more' ... Even though best practices aren't perfect for everyone, if I see a repeat of a

tip between the two [Guidedogs], then it’s kind of like two different people gave you [the same piece of] feedback. So that seems like it’s elevated as more important.”

**5.1.3 Optimized and Selective Tool Usage.** Given the high volume of feedback output by the five different personas, participants often had to develop strategies to use the Guidedogs tool more effectively and efficiently. For example, some initially read the feedback from all Guidedogs to understand their perspective before choosing a small set of personas to consult in future iterations. P2 shared their reasoning for picking two specific Guidedogs for use during the description workshop: “*by the end of the three days, [I chose] two — these are the two that I’m going to concentrate on. ... It’s a time restriction... so I don’t have time to read all of [the other Guidedogs]... I’m going to marry these two together and make them work.*”

Others chose to sample feedback from the Guidedogs on a case-by-case basis. Participants characterized the level of feedback they received from the Guidedogs as a “*Goldilocks point*” (P6) once they figured out how to parse it: “*it wasn’t overbearing, it wasn’t too much, it was just right*” (P12). P1 also mentioned that users could have their “*favorite Guidedog for a certain situation,*” alluding to how individual Guidedog personas could be more or less useful depending on the image type. P14 acknowledged that the amount of feedback could be overwhelming, but after some usage, it was largely manageable: “*I didn’t feel pressured to read all of it... I’m already just jumping around to see what I want.*” However, she also acknowledged that her strategy for optimizing her Guidedogs tool use could be due to her prior experience with writing AD, and she wondered how novices would have reacted to the quantity of feedback. Similarly, P1 also selected specific feedback from the tool, and drew an analogy between consulting the panel of Guidedogs and discussing with a set of trusted people:

“To me the value is... you are the head of a company or something, and you [have] four different people on your staff, or you’re the head of a country, and you have your cabinet. ... And then you’re getting these different perspectives. You might take a little from one or a little from another... in the role of the arbiter... until you feel more confident in making your own judgments about what you want.”

In addition to gathering feedback from a variety of perspectives, participants used the Guidedogs to aid with a specific step of the writing process. They used the Guidedogs as a way to double check instincts or as a sounding board, acknowledging that “*we would have to make a judgment call at the end of the day, but that’s the way it should be*” (P13). Some also found the tool helpful for copyediting and polishing final drafts, or for “*enhanc[ing] what we wrote there if we weren’t totally satisfied with it*” (P13).

**5.1.4 Lasting Impact.** Participants expressed that the Guidedogs could be helpful both during the workshop and beyond. For example, the learnings that they gained from using the tool early on in the description workshop could be applicable when writing descriptions later on. In particular, “*high impact suggestions*” (P11), such as reminders about including spatial information, would stick with the describers: “*we were able to integrate that across the entire rest of our descriptions*” (P11). Similarly, P6 elaborated on the importance of educating describers for long-term improvement: “*that learning comes out and [describers] start applying it to everything that they do; they’ve got that mental checklist and a little Guidedog voice on their shoulder*” (P6). As a professional audio describer, P2 also mentioned that her practice of receiving detailed and critical feedback from the Guidedogs influenced her current description practices with coworkers: “*I feel like I do force my coworkers to read my descriptions more now, and not just read them and say, ‘Hey, what do you think about this?’ But I really adamantly say, ‘No, what exactly do you want to know?’*”

Some participants also expressed that they appreciated how the tool was fun, easy, and convenient to use. P6 highlighted that using the Guidedogs felt like *“a form of play where, at the same time, you’re learning and then [you] bring that to the next session in the evening.”*

Overall, participants found the Guidedogs tool helpful for improving descriptions and realistic to the experience of catering descriptions to a wide variety of audiences, with concrete takeaways that improved both novices’ and experts’ description-writing skills long-term.

## 5.2 Social Affordances

The Guidedogs were helpful for collaborating within the description workshop teams. For novices, the Guidedogs provided an affordance of seeking feedback independently and increasing their sense of privacy and safety when writing descriptions for the first time. Among teams, the Guidedogs could also spark conversations and resolve conflicts.

**5.2.1 Seeking Feedback Independently.** While AD writers often seek feedback from blind and low vision end users to improve their AD drafts, the Guidedogs allowed participants to obtain simulations of this feedback more immediately and conveniently than they could by asking other team members (either sighted or BLV). Though he did not use the feature during the description workshop, P8 thought that using the Guidedogs could simulate *“a good testing session”* and could unblock anybody without access to end users, but acknowledged that *“it would still be good to get actual people at some point.”*

Among participants with relatively more experience with description-writing, they thought that this affordance of the Guidedogs was similar to their experiences outside of the workshop of writing and iterating on descriptions to cater to diverse perspectives. When testing out the Guidedogs prior to the description workshop, P8 shared: *“it’s interesting to see when the Guidedogs conflict... I mean, this is [like] real life... I don’t think it would be as helpful if it was just one Guidedog.”* Similarly, as someone who had participated in previous iterations of the description workshop without the Guidedogs, P11 felt that this experience of working with the tool *“seemed very true to the experiences I’ve had with other [description workshops] and working with a variety of... different people’s preferences.”* P13 also remarked on the quality of feedback from the Guidedogs: *“[it] provided some fairly insightful perspectives I wouldn’t necessarily expect from an AI tool... it really was feedback that I could have gotten from another human being.”*

**5.2.2 Increasing a Sense of Privacy and Safety.** Other valuable affordances of the Guidedogs included helping participants build their confidence and have greater privacy when seeking feedback and even learning more about AD and the BLV community. For example, P7 explained how novices on their team made use of the Guidedogs feature as a way to learn more about AD and ramp up to writing. While a small set of volunteers during the description workshop were trained professional audio describers, a vast majority were novice describers who had relatively limited knowledge and exposure to accessibility. As a result, participants sometimes did a first pass for feedback with the Guidedogs as a quick confidence check. P7 shared her thoughts about this use case:

*“There’s a lot of, I don’t want to say shyness, but I feel like there’s a lot of hesitation to put [your description] out there when your team consists of people who have done this before, maybe do this for a living. Maybe [they] are users of this thing. And people don’t want to say something wrong, which is valid. It’s the human experience, but I think [the Guidedogs tool] provides that safe space where they can get a little bit more feedback before they bring it to the group.”*

For novices, the automated nature of the Guidedogs meant that they “*offered a sense of privacy and safety when you were using them*” (P6). P6, who identifies as DeafBlind, highlighted the positive nature of her teammates’ experiences with the Guidedogs:

“Those dogs are much easier to listen to when they’re politely telling you that your description of a collage is actually quite rubbish and nobody would understand what you said... rather than feel you’ve let yourself down in front of another human.”

**5.2.3 Navigating Time Pressures and Team Conversations.** Multiple participants shared that the description workshop had a heavy workload and could be difficult to manage without the right techniques. Some mentioned that this sense of time pressure prevented them from using the Guidedogs tool, with P7 acknowledging, “*genuinely, with the workload I had during that time, I just didn’t have time to.*” To navigate this, during the description workshop, some teams took a divide and conquer approach of “*splitting up descriptions*” (P2) to ensure that they could describe all images required during the short timeframe of the workshop.

Most teams convened about the descriptions prior to finalizing them, and some found that the feedback that the Guidedogs provided could support team conversations. For example, P1 shared that the Guidedogs were helpful during disagreements, as they could expand people’s perspectives and make them more open to different opinions: “*by talking it through and saying what we disagreed about... that’s exactly what is the value of it, that it challenges you to think, if you’re open... I want to hear what other people think.*” P17 also felt that the Guidedogs helped to “*get some conversations going... it provided specific examples of different audiences we are trying to reach... I think that really led to discussion among the team.*” Among P13’s team, the Guidedogs were also used as a conversation starter: “*it provided us enough detailed information that we could sit in a room and have a conversation and say, ‘Well, Playful Pearl says this, what do you guys think?’*” Similarly, P8 felt that the Guidedogs provided a starting point for critiquing teammates’ descriptions in a constructive way: “*it starts the conversation more effectively when you’re reviewing it... what I like about it is that [the Guidedogs are] not sugar coating it.*”

### 5.3 Constraints of the Guidedogs Tool

While most of the participants who used the Guidedogs had positive experiences, some shared that the tool could sometimes be more confusing than helpful. They also described key situations when the amount of feedback received from the Guidedogs negatively affected their experience.

**5.3.1 Causing Confusion or Disappointment for Novices.** Participants acknowledged that novice describers might not have strong intuitions for what constitutes high-quality AD, which could make diverse feedback streams confusing. For example, P1 mentioned: “*if you’re really early on, and you kind of don’t even understand what [the] Guidedogs are... hearing the opinion of different people can be more confusing than it can be helpful.*” They felt that “*the annoyance factor was directly proportional to where I was at in the draft*” (P1) as they had difficulty determining which pieces of advice to implement at the beginning of the drafting process. Similarly, some participants felt some confusion about “*which voice we should be going for*” (P17). Others described that they were confused initially before figuring out how to best integrate the Guidedogs tool in their workflow. For example, throughout the three-day description workshop, P2 was able to navigate the “*learning curve [and built] a little bit of understanding how to make it work best to your writing style.*” P2 shared that they ended up using the Guidedogs “*on everything from the very first description to the end [and found] more success the more [they] used them.*”

Given that the Guidedogs were often used to validate draft descriptions, participants were sometimes disappointed when they received constructive feedback on descriptions that they thought were already well done. In particular, P6 described the experiences that her team had when seeking feedback from the tool:

“You put your hard work into that Guidedog and you want it not to change anything... it’s almost like getting an A grade from your teacher. When the Guidedog comes back and goes, ‘Well, it was pretty good, but I also want to know blah, blah, blah,’ it’s quite disheartening.”

Participants who did not use the Guidedogs shared some sources of confusion that impacted their degree of engagement. P5 mentioned that the interface was confusing to her and she had difficulty finding the tool. Instead, she opted to use external AI assistants that she already had experience using, such as Gemini and Hemingway, to copy edit and support her description writing process. Some also discussed how the speed and workload of the description workshop made it more difficult to engage with a new tool, even if it was helpful. For example, P16 shared: *“recognizing that the workload is pretty heavy, and you’re rushing at some points... my tired brain after a certain point was like, ‘Okay, I can’t filter between the Guidedogs information and my information.’”*

**5.3.2 Overwhelming Amount of Feedback.** The primary piece of constructive feedback that participants shared about the Guidedogs was that there was too much feedback, to the point of being overwhelming and unhelpful. While most participants who used the tool rated the amount of feedback to be a good middle ground between not enough and too much, three participants (P2, P11, and P16) rated the level of feedback as being more than was necessary or helpful. P11 felt that the Guidedogs provided some actionable and high impact suggestions, but could be buried under less useful information: *“not every Guidedog provided too much feedback, but [with] the number of Guidedogs plus the amount of feedback that some were giving, it was a lot to digest in one go.”* Even though participants developed strategies to parse through the Guidedogs, she described that it could be *“kind of hard to thread the needle with all of that feedback coming through... I found myself scanning the first few blurbs... and then kind of losing a little steam as I went down”* (P11). Others, such as P12, also felt that there was more feedback than needed: *“I think we were just pleased with having two or three helpful dogs.”*

Both P2 and P16 encountered situations where the Guidedogs could lead them *“down the rabbit hole”* (P2) and cause them to get *“stuck in a feedback loop”* (P16) after multiple iterations. In her attempt to optimize descriptions to *“appease”* (P2) all of the Guidedogs, P2 shared: *“I wrote this perfect description for this [Guidedog], but now it’s not the perfect description for this one. So now how do I make it perfect for this? [I was] just circling around and around and around.”* From her repeated use of the Guidedogs on one description, she realized that they would always provide feedback even if earlier feedback had already been taken into account: *“it will never be perfect for them, apparently. They will... keep telling you to change things”* (P2). P16 also found the Guidedogs to be helpful to a certain point, before the feedback became overwhelming. She mentioned that *“folks had a similar experience... don’t put all your eggs in the Guidedog basket. They’re super helpful, but if you run the same content through, over and over again, you’re going to hit a wall at a certain point”* (P16). Overall, while the Guidedogs provided useful insights, there were limits to how helpful the tool could be before becoming overwhelming or unproductive.

## 5.4 Thoughts on the Guide Dog Metaphor

Almost all participants 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).

Participants also shared their thoughts about the specific personalities assigned to the individual Guidedogs. In particular, Temperamental Teena elicited the most reactions. She was helpful for some — “the teenager voice particularly made a difference” (P6) and her feedback reminded describers that “this [description] is so boring, come on, you can do better” (P16). However, she was also “off-putting [or] abrasive” (P5) for others. P2 mentioned that Teena’s persona was one that she “just stopped reading... [it was] more entertaining than anything else.” Describers generally appreciated Focused Frankie’s emphasis on conciseness, and also mentioned that Playful Pearl’s strength was that she reinforced “the idea that we need to make the descriptions fun and not so serious... that people want to enjoy reading the descriptions” (P17).

## 5.5 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 regarding ethics, quality, and accuracy.

Some participants mentioned that AI tools could help them describe more quickly and creatively, without compromising a description’s overall quality. For example, though he did not use the tool during the description workshop, 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.”

Participants also explored the nuances surrounding AI usage, and shared how they reacted to the rapidly developing AI landscape. P2 acknowledged that some AI tools could be helpful and should be explored, but also shared concerns: “there’s a fine line between just blowing it off because it’s AI and then overusing it.” Similarly, others felt that “there are certainly ways it can be abused... as with all technology, but I think that the ways it can actually help and serve are what’s

really exciting” (P1). In terms of accuracy, P10 felt that *“it didn’t make any mistakes from what I could tell [but] I guess it depends on what you define as a mistake... [it made] a certain number of what I would describe as generalizations or perhaps assumptions.”* He emphasized that more development and testing of AI systems as necessary before phasing out human involvement. P9, a blind participant who was enthusiastic about AI, shared that some AI-generated descriptions from off-the-shelf applications such as Be My AI were helpful in her day-to-day life. She sometimes disagreed with other BLV community members about who or what could be involved in AD: *“[they think] everything must be objective all the time... [and] it has to be an expert to describe this that has a wealth of knowledge about it, which is unrealistic”* (P9).

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 AI and Apple’s Siri for personal and experimental purposes. However, she felt that it was inappropriate to use AI during the description workshop, a stance influenced by her role as a Justice (a paid professional providing feedback on novice-written descriptions). Out of concerns regarding professionalism 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.

## 6 Discussion

### 6.1 Engaging Novices in Access Efforts

Given the importance and prevalence 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., [9, 27]).

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 scale AD 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 [17, 29, 30].

### 6.2 Generative AI and Creativity

With the advancement of GenAI systems, such as ChatGPT-4o [49] and Claude 3.7 [4], computers now have the capability to replicate many aspects of human creative work and other tasks that humans find meaningful and valuable. In the context of the description workshop and the Guidedogs tool, participants generally identified valuable affordances specific to GenAI integration, such as rapid iteration without burdening BLV community members and increasing novices’ sense of privacy through avoiding embarrassment in front of team members. However, some of the blind and low vision participants that we interviewed shared their hesitancy about allowing AI to replace human describers, especially in this formal context in which descriptions were to be widely distributed by the National Park Service.

The human-computer interaction research community has grappled with this question about how to best integrate GenAI systems into creative contexts. For example, Kawakami and Venkatgiri highlight the tension between using GenAI to augment artists’ expression while these systems simultaneously scrape their data and replace valued human labor and creativity [31]. GenAI assistants have also been used to write poetry and create other forms of “*noncommercial creative expression*” [21]; however, these usages are often met with controversy and dissatisfaction [32]. Audio descriptions are critical for “*adding artistic and rhetorical value, not simply transposing or distilling meanings*” ([19], p. 140). Many researchers have previously advocated for human-in-the-loop involvement as they found AI-supported AD systems could not independently create sufficient and serviceable descriptions (e.g., [37, 44, 62, 63]). While GenAI systems can now create facsimiles of these artistic expressions, the process of writing AD is equally as meaningful as the final output. We advocate for humans to remain involved in the AD creation process and recommend designing GenAI systems to support human efforts, rather than entirely replace humans in the name of speed and scale.

There are many benefits to maintaining human involvement while integrating AI support. In terms of benefits to creators, this provides a creative outlet for humans to engage in a group activity and exercise their writing skills. In terms of benefits to descriptions, human-in-the-loop approaches ensure that humans have the opportunity to make artistic and poetic choices to humanize descriptions and correct errors or hallucinations generated by AI systems, improving the overall quality of the resulting AD. Regarding the benefits to the accessibility ecosystem, this also allows human creators (both novices and experts) more opportunities to become involved in accessibility efforts and learn about the importance of increasing access to visual materials.

### 6.3 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 have low vision have started to embrace AI tools in their daily lives (e.g., [1, 26]), often for increasing agency in accessing visual descriptions. However, when using AI to create official and published work (such as audio described NPS brochures), there is still a pressing need for greater accuracy, verifiability, and trust prior to more widespread adoption of these AI tools.

The Guidedogs tool was also useful for navigating and mediating collaboration during the description workshop. Some off-the-shelf AI tools have specific versions designed for collaborative scenarios (e.g., ChatGPT Personal compared to their Team and Enterprise versions [47]). In our case, while the tool was designed for individual use, we found that the composition of participants and the context of the description workshop led to unanticipated positive affordances of the Guidedogs tool among groups. For example, some novice volunteers used the Guidedogs as a safe way to conduct a confidence check, allowing them to iterate on their draft descriptions before sharing them with the broader team. Others found the Guidedogs helpful for gathering feedback in a much more rapid fashion. While AI-based tools cannot replace human team members and invaluable feedback from blind and low vision people with lived experience, we recommend that technologists consider how description-writing tools designed for individual use can support social dynamics and mixed-ability collaboration.

Extending prior work on writing tools (e.g., [35]) and providing feedback for image describers (e.g., [37, 41, 64]), 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 [38]. 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. Furthermore, our metaphor of guide dogs aligns with this recommendation: while real-life guide dogs provide valuable guidance to blind and low vision people, humans ultimately retain agency over their decisions. Lastly, we encourage AI systems to **leverage familiar metaphors and designs to support user adoption**, given participants' positive responses to the Guidedogs design of our AI tool.

#### 6.4 Limitations and Future Work

There are some limitations to our system. Notably, the Guidedogs system only processed text input and output and 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. This also meant that the tool could not serve as a way for blind or low vision description writers to identify or confirm visual details, which is something that often occurs when receiving descriptions from trusted family and friends. Additionally, our system featured five personas predetermined by the research team, based on the context of describing National Park Service brochures. Future work could design additional Guidedog personas to explore how GenAI systems could support description-writing across different contexts (e.g., live theater, STEM classes, sporting events, etc.).

Our interview study also has several limitations. 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 many BLV people with direct experience using the Guidedogs tool. While some BLV interview participants were familiar with the tool, one specifically mentioned ethical concerns with using the AI tool given that she was being paid for her expertise. As a result, we did not receive feedback from BLV people who used the Guidedogs for writing descriptions themselves. Furthermore, all participants were given description output from all of the Guidedogs at once. Future iterations of the tool could allow users to select which specific Guidedog personas from which they are seeking feedback, which could also provide more detailed in-situ insights about which personas were helpful for which types of descriptions. 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.

## 7 Conclusion

Given the widespread nature of visual media, it is important to increase both the quality and quantity of descriptions available to blind and low vision audiences. To work towards this goal, we designed the Guidedogs tool to provide diverse and immediate feedback to novice describers' draft audio descriptions created during a multi-day description workshop. Through semi-structured interviews with workshop participants, we found that describers appreciated the design of the tool and found it useful for seeking inspiration and exploring different perspectives through feedback from five personas. However, it is also critical for the research community to consider the ethical consequences of integrating GenAI into accessibility work, and we advocate for continuing human-in-the-loop processes to ensure quality and increase awareness about the importance of accessibility. Through this work, we contribute design rationale and insights regarding the usefulness, affordances, and constraints of generative AI description writing assistants.

## References

- [1] Rudaiba Admin 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] Aira. 2025. *Aira*. <https://aira.io/>
- [3] Hussam Alkaissi and Samy I McFarlane. 2023. Artificial hallucinations in ChatGPT: implications in scientific writing. *Cureus* 15, 2 (2023). <https://doi.org/10.7759/cureus.35179>
- [4] Anthropic. 2025. *Claude 3.7 Sonnet and Claude Code*. <https://www.anthropic.com/news/claude-3-7-sonnet>
- [5] Be My Eyes. [n. d.]. *Announcing ‘Be My AI,’ Soon Available for Hundreds of Thousands of Be My Eyes Users*. <https://www.bemyeyes.com/blog/announcing-be-my-ai>
- [6] Karim Benharrak, Tim Zindulka, Florian Lehmann, Hendrik Heuer, and Daniel Buschek. 2024. Writer-defined AI personas for on-demand feedback generation. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–18. <https://doi.org/10.1145/3613904.3642406>
- [7] 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*. 1–19. <https://doi.org/10.1145/3411764.3445498>
- [8] 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>
- [9] 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 23rd annual ACM symposium on User interface software and technology*. 333–342. <https://doi.org/10.1145/1866029.1866080>
- [10] Jeffrey P Bigham, Richard E Ladner, and Yevgen Borodin. 2011. The design of human-powered access technology. In *The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility*. 3–10. <https://doi.org/10.1145/2049536.2049540>
- [11] 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>
- [12] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101. <https://doi.org/10.1191/1478088706qp0630a>
- [13] Joy Buolamwini and Timnit Gebru. 2018. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency*. PMLR, 77–91. <https://proceedings.mlr.press/v81/buolamwini18a.html>
- [14] Michele A Burton, Erin Brady, Robin Brewer, Callie Neylan, Jeffrey P Bigham, and Amy Hurst. 2012. Crowdsourcing subjective fashion advice using VizWiz: challenges and opportunities. In *Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility*. 135–142. <https://doi.org/10.1145/2384916.2384941>
- [15] Tuhin Chakrabarty, Vishakh Padmakumar, Faeze Brahman, and Smaranda Muresan. 2024. Creativity Support in the Age of Large Language Models: An Empirical Study Involving Professional Writers. In *Proceedings of the 16th Conference on Creativity & Cognition*. 132–155. <https://doi.org/10.1145/3635636.3656201>
- [16] 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>
- [17] 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>
- [18] Victoria Clarke and Virginia Braun. 2017. Thematic analysis. *The journal of positive psychology* 12, 3 (2017), 297–298. <https://doi.org/10.1080/17439760.2016.1262613>
- [19] Jay Timothy Dolmage. 2014. *Disability rhetoric*. Syracuse University Press.
- [20] 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>
- [21] Michele Elam. 2023. Poetry will not optimize; or, what is literature to AI? *American literature* 95, 2 (2023), 281–303. <https://doi.org/10.1215/00029831-10575077>
- [22] Envision. [n. d.]. *Envision - Perceive Possibility*. <https://www.letsenvision.com/>
- [23] Katy Ilonka Gero, Meera Desai, Carly Schnitzler, Nayun Eom, Jack Cushman, and Elena L Glassman. 2024. Creative Writers’ Attitudes on Writing as Training Data for Large Language Models. *arXiv preprint arXiv:2409.14281* (2024). <https://arxiv.org/pdf/2409.14281>
- [24] 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>
- [25] Cole Gleason, Amy Pavel, Emma McCamey, Christina Low, Patrick Carrington, Kris M Kitani, and Jeffrey P Bigham. 2020. Twitter A11y: A browser extension to make Twitter images accessible. In *Proceedings of the 2020 chi conference on human factors in computing systems*. 1–12. <https://doi.org/10.1145/3313831.3376728>
- [26] 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/10.1145/3654777.3676375>



- //doi.org/10.1145/3613904.3642211
- [27] 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>
- [28] 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>
- [29] 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>
- [30] 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>
- [31] Reishiro Kawakami and Sukrit Venkatagiri. 2024. The Impact of Generative AI on Artists. In *Proceedings of the 16th Conference on Creativity & Cognition*. 79–82. <https://doi.org/10.1145/3635636.3664263>
- [32] Madison Malone Kircher. 2025. People Love Studio Ghibli. But Should They Be Able to Recreate It? <https://www.nytimes.com/2025/03/27/style/ai-chatgpt-studio-ghibli.html>
- [33] Elizabeth Kupferstein, Yuhang Zhao, Shiri Azenkot, and Hathaitorn Rojnirun. 2020. Understanding the use of artificial intelligence based visual aids for people with visual impairments. *Investigative Ophthalmology & Visual Science* 61, 7 (2020), 932–932. <https://iovs.arvojournals.org/article.aspx?articleid=2766798>
- [34] Philippe Laban, Jesse Vig, Marti Hearst, Caiming Xiong, and Chien-Sheng Wu. 2024. Beyond the chat: Executable and verifiable text-editing with llms. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–23. <https://doi.org/10.1145/3654777.3676419>
- [35] 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>
- [36] Yiren Liu, Pranav Sharma, Mehul Jitendra Oswal, Haijun Xia, and Yun Huang. 2024. Personaflow: Boosting research ideation with llm-simulated expert personas. *arXiv preprint arXiv:2409.12538* (2024). <https://arxiv.org/abs/2409.12538>
- [37] Kelly Avery 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>
- [38] 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. <https://doi.org/10.1145/3696449>
- [39] Microsoft. [n. d.]. *Seeing AI*. <https://www.seeingai.com/>
- [40] Midjourney. [n. d.]. *Midjourney*. <https://www.midjourney.com/home>
- [41] 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>
- [42] Mycroft AI. [n. d.]. *Mimic TTS*. <https://mycroft-ai.gitbook.io/docs/mycroft-technologies/mimic-tts>
- [43] 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. <https://doi.org/10.1145/3663548.3675617>
- [44] 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>
- [45] National Park Service. 2019. *Accessibility and Audiovisual Media FAQ*. <https://www.nps.gov/subjects/hfc/accessibility-and-audiovisual-media-faq.htm>
- [46] National Park Service. 2025. *Accessibility (U.S. National Park Service)*. <https://www.nps.gov/subjects/accessibility/index.htm>
- [47] Open AI. [n. d.]. *ChatGPT Pricing*. <https://openai.com/chatgpt/pricing/>
- [48] Open AI. [n. d.]. *GPT-3.5 Turbo*. <https://platform.openai.com/docs/models/gpt-3.5-turbo>
- [49] Open AI. 2024. *Hello GPT-4o*. <https://openai.com/index/hello-gpt-4o/>
- [50] Renée Otmar, Rose Michael, Sharon Mullins, and Katherine Day. 2024. Ethics and the use of generative AI in professional editing. *AI and Ethics* (2024), 1–13. <https://doi.org/10.1007/s43681-024-00521-7>
- [51] Daisuke Sato, Masatomo Kobayashi, Hironobu Takagi, and Chieko Asakawa. 2010. Social accessibility: the challenge of improving web accessibility through collaboration. In *Proceedings of the 2010 International Cross Disciplinary Conference on Web Accessibility (W4A)*. 1–2. <https://doi.org/10.1145/1805986.1806024>
- [52] 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>
- [53] 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>



- [54] Social Audio Description Collective. [n. d.]. *Our Values and Approach*. <https://socialaudiodescription.com/about/>
- [55] Abigale Stangl, Meredith Ringel Morris, and Danna Gurari. 2020. "Person, Shoes, Tree. Is the Person Naked?" What People with Vision Impairments Want in Image Descriptions. In *Proceedings of the 2020 chi conference on human factors in computing systems*. 1–13. <https://doi.org/10.1145/3313831.3376404>
- [56] Hironobu Takagi, Susumu Harada, Daisuke Sato, and Chieko Asakawa. 2013. Lessons learned from crowd accessibility services. In *Human-Computer Interaction-INTERACT 2013: 14th IFIP TC 13 International Conference, Cape Town, South Africa, September 2-6, 2013, Proceedings, Part I* 14. Springer, 587–604. [https://doi.org/10.1007/978-3-642-40483-2\\_42](https://doi.org/10.1007/978-3-642-40483-2_42)
- [57] 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*. 1–17. <https://doi.org/10.1145/3613904.3642839>
- [58] 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>
- [59] Meredith Whittaker, Meryl Alper, Cynthia L Bennett, Sara Hendren, Liz Kaziunas, Mara Mills, Meredith Ringel Morris, Joy Rankin, Emily Rogers, Marcel Salas, et al. 2019. Disability, bias, and AI. *AI Now Institute* 8 (2019), 11. <https://www.academia.edu/download/95505126/disabilitybiasai-2019.pdf>
- [60] 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>
- [61] Catherine Yeh, Gonzalo Ramos, Rachel Ng, Andy Huntington, and Richard Banks. 2025. Ghostwriter: Augmenting collaborative human-ai writing experiences through personalization and agency. *arXiv preprint arXiv:2402.08855* (2025). <https://arxiv.org/abs/2402.08855>
- [62] Beste F Yuksel, Pooyan Fazli, Umang Mathur, Vaishali Bisht, Soo Jung Kim, Joshua Junhee Lee, Seung Jung Jin, Yue-Ting Siu, Joshua A Miele, and Ilmi Yoon. 2020. Human-in-the-loop machine learning to increase video accessibility for visually impaired and blind users. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference*. 47–60. <https://doi.org/10.1145/3357236.3395433>
- [63] Beste F Yuksel, Soo Jung Kim, Seung Jung Jin, Joshua Junhee Lee, Pooyan Fazli, Umang Mathur, Vaishali Bisht, Ilmi Yoon, Yue-Ting Siu, and Joshua A Miele. 2020. Increasing video accessibility for visually impaired users with human-in-the-loop machine learning. In *Extended abstracts of the 2020 CHI conference on human factors in computing systems*. 1–9. <https://doi.org/10.1145/3334480.3382821>
- [64] 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>

## A Guidedogs Feedback

Table 2. Examples of feedback provided by two Guidedogs, Temperamental Teena and Focused Frankie.

### Draft Description

Side one provides an overview of the two desert ecosystems found in the park, the Mojave and Colorado deserts. On the left third of the page, a background photo shows a typical Mojave desert landscape with a Joshua tree in the foreground and large boulders behind. Overlaid on this background image, a column of small circular inset photos arranged vertically show some of the wildlife residents of the park. In the center third of the page, text and collage illustrations of the Mojave Desert and the Colorado Desert detail the plants and animals specific to each ecosystem. Below these collages is a brief description of the geologic processes that form the park's iconic boulders. On the right third of the brochure, a background image shows a typical Colorado desert landscape. At the top right corner, a small circular inset photo of an oasis is paired with text.

### Feedback

**Temperamental Teena:** Hey there! So, I checked out your description, and I gotta say, it's not bad, but it could definitely use a little more pizzazz to grab my attention. Let's break it down and see how we can make it more exciting and interesting, shall we?

First off, you're talking about two desert ecosystems, the Mojave and Colorado deserts. That's cool and all, but let's spice it up a bit. How about we dive into the cultural significance of these deserts? Tell me about the Native American tribes that have called these deserts home for centuries. Maybe mention some of the unique traditions or stories that have been passed down through generations. That'll really make me sit up and take notice!

Now, let's talk visuals. You've got these background photos and inset images, which is a good start. But let's make it more engaging. Instead of just saying "a typical Mojave desert landscape," describe the vibrant colors of the sand and the rugged beauty of the Joshua tree. Paint a picture with your words, you know? And when you mention the wildlife residents, give me some fun facts about them! Tell me about the quirky behaviors or adaptations that make them stand out. [...]

So, to sum it up, spice up your description with cultural tidbits, use vivid language to bring the visuals to life, and sprinkle in some fun facts and scientific explanations. Remember, Teena wants to be entertained and engaged, so make it exciting and interesting! You got this!

**Focused Frankie:** Thank you for sharing your description with me. I can tell you are putting a lot of effort into providing a comprehensive overview of the desert ecosystems in the park. Overall, your description effectively conveys the layout and content of the brochure. However, I have a few suggestions to enhance the clarity and engagement of your description.

1. Begin with a captivating introduction: To grab the listener's attention, consider starting with a captivating sentence that highlights the unique features or beauty of the desert ecosystems. For example, you could say, "Welcome to the breathtaking world of the park's desert ecosystems, where the Mojave and Colorado deserts converge in a harmonious dance of rugged landscapes and vibrant wildlife."

2. Provide more vivid and sensory language: To paint a more vivid picture in the listener's mind, incorporate sensory language that appeals to their senses. Instead of simply stating "a typical Mojave desert landscape," you could say, "Immerse yourself in the vast expanse of the Mojave desert, where golden sand dunes stretch as far as the eye can see, and the iconic Joshua trees stand tall, casting intriguing shadows on the arid ground." [...]

Remember, Focused Frankie appreciates a balance between informative details and engaging storytelling. By incorporating these suggestions, your audio description will not only provide valuable information but also captivate and inspire the listener to explore the wonders of the park's desert ecosystems. Well done on your efforts so far, and keep up the great work!