Memory Reviver: Supporting Photo-Collection Reminiscence for People with Visual Impairment via a Proactive Chatbot

Shuchang Xu
Hong Kong University of Science and Technology
Hong Kong, China
xschci@gmail.com

Chang Chen
Hong Kong University of Science and Technology
Hong Kong, China
cchenda@connect.ust.hk

Zichen Liu
Hong Kong University of Science and Technology
Hong Kong, China
zliucz@connect.ust.hk

Xiaofu Jin
Hong Kong University of Science and Technology
Hong Kong, China
xjinao@connect.ust.hk

Lin-Ping Yuan*
Hong Kong University of Science and Technology
Hong Kong, China
lyuanaa@cse.ust.hk

Yukang Yan
University of Rochester
New York, United States
yukang.yan@rochester.edu

Huamin Qu
Hong Kong University of Science and Technology
Hong Kong, China
huamin@cse.ust.hk

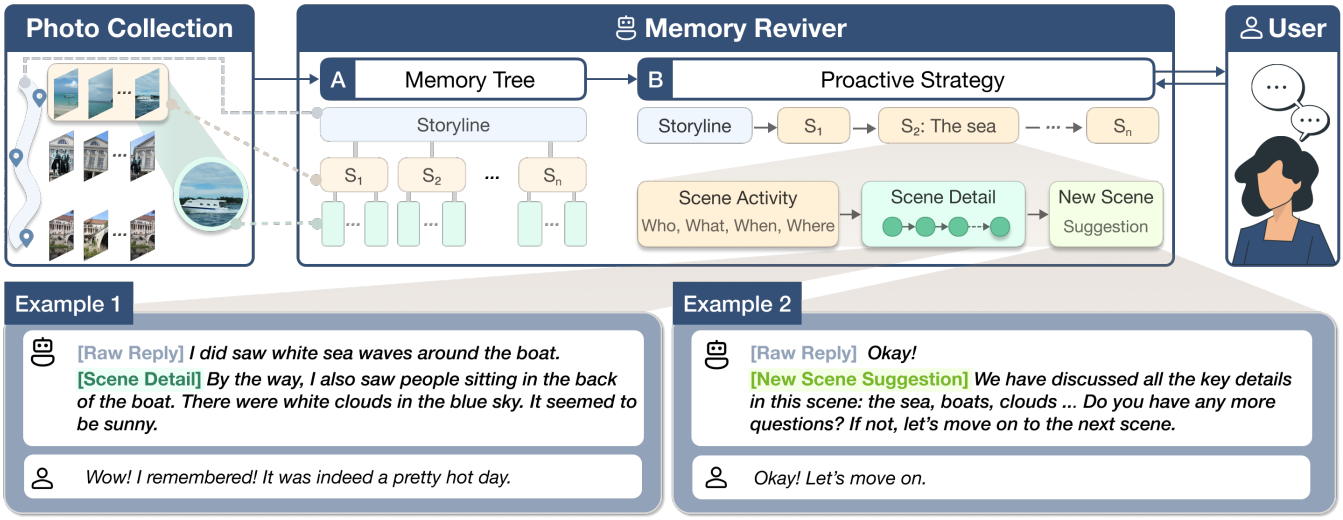

Figure 1: Memory Reviver is a proactive chatbot that actively guides people with visual impairment (PVI) to reminisce with a photo collection. It incorporates two novel features: (A) a *Memory Tree*, which uses a hierarchical structure to organize information in a photo collection; and (B) a *Proactive Strategy*, which actively delivers information to users at proper conversation rounds. Powered by the two features, Memory Reviver begins the chat with a clear storyline, helps users recall past activities, enriches their memories by gradually presenting details (Example 1), and suggests new scenes at proper rounds (Example 2).

*Corresponding author.

UIST '24, October 13–16, 2024, Pittsburgh, PA, USA

ABSTRACT

Reminiscing with photo collections offers significant psychological benefits but poses challenges for people with visual impairment (PVI). Their current reliance on sighted help restricts the flexibility

ACM ISBN 979-8-4007-0628-8/24/10
https://doi.org/10.1145/3654777.3676336

of this activity. In response, we explored using a chatbot in a preliminary study. We identified two primary challenges that hinder effective reminiscence with a chatbot: the scattering of information and a lack of proactive guidance. To address these limitations, we present Memory Reviver, a proactive chatbot that helps PVI reminisce with a photo collection through natural language communication. Memory Reviver incorporates two novel features: (1) a *Memory Tree*, which uses a hierarchical structure to organize the information in a photo collection; and (2) a *Proactive Strategy*, which actively delivers information to users at proper conversation rounds. Evaluation with twelve PVI demonstrated that Memory Reviver effectively facilitated engaging reminiscence, enhanced understanding of photo collections, and delivered natural conversational experiences. Based on our findings, we distill implications for supporting photo reminiscence and designing chatbots for PVI.

CCS CONCEPTS

• **Human-centered computing → Accessibility systems and tools**; *Empirical studies in accessibility*.

KEYWORDS

Visual Impairment, Blind, Low Vision, Photo, Collection, Reminiscence, Memories, Chatbot, Conversational Agent

ACM Reference Format:
Shuchang Xu, Chang Chen, Zichen Liu, Xiaofu Jin, Lin-Ping Yuan, Yukang Yan, and Huamin Qu. 2024. Memory Reviver: Supporting Photo-Collection Reminiscence for People with Visual Impairment via a Proactive Chatbot. In *The 37th Annual ACM Symposium on User Interface Software and Technology (UIST '24), October 13–16, 2024, Pittsburgh, PA, USA.* ACM, New York, NY, USA, 17 pages. https://doi.org/10.1145/3654777.3676336

1 INTRODUCTION

Reminiscence, the activity of recalling life experiences from one's past [71, 73], has been shown to improve mental well-being [14, 47], foster social connections [3], and promote personal growth [22, 26]. This process allows individuals to reflect on memorable moments and milestones, often by browsing through photo collections [5, 17, 48]. However, reminiscing with a photo collection is challenging for people with visual impairment (PVI). Although tools have been explored to help PVI examine individual photos [37, 51, 61, 62], they lack features for a comprehensive reminiscing experience, such as providing a holistic overview of the photo collection. Consequently, PVI often rely on sighted people to reminisce with photo collections [37, 82]. This dependence limits the frequency of their reminiscing activities and impedes them from deriving the associated psychological benefits [37, 82].

To address this limitation, we propose designing a chatbot that enables PVI to reminisce with photo collections through natural language communication. This chat-based approach is already familiar to PVI through their interactions with sighted helpers [7], rendering it a suitable alternative when sighted help is unavailable or not preferred [37]. Moreover, chatbots have been shown to effectively support emotional communication [60, 72], making them a promising tool for facilitating reminiscence.

To inform the chatbot design, we conducted a formative study with eight PVI. They were invited to reminisce with their photo collections by conversing with a naïve chatbot based on GPT-4V [81], the state-of-the-art large multimodal model. Our research uncovered significant shortcomings in directly utilizing GPT-4V for reminiscence activities among PVI, primarily due to disorganized conversation flow and the chatbot's non-proactive interaction style. Firstly, the "one question, one answer" communication style led to the information being scattered across multiple rounds. This made it hard for participants to recall and organize details, preventing them from forming a clear story about their past. Secondly, since participants could not visually explore new scenes, the chatbot's lack of proactive guidance further challenged participants to engage deeply in the conversation.

To address these challenges, we present Memory Reviver, a proactive chatbot that actively guides PVI to reminisce with a photo collection. Memory Reviver is tailored for photo collections about a specific event. It incorporates two novel features: (1) an information architecture of ***Memory Tree***, which extracts the information in a photo collection into a hierarchical structure; and (2) a ***Proactive Strategy***, which actively delivers information to users at recognized proper conversation rounds. Powered by the two features, Memory Reviver delivers a natural conversation flow. It begins the conversation with a clear storyline, helps users recall past activities, and enriches their memories by gradually presenting details. After thoroughly examining the details of a specific scene, Memory Reviver proactively suggests new scenes. This seamless blend of guided exploration with free-form dialogue allows Memory Reviver to offer an experience that closely resembles natural reminiscence.

To evaluate Memory Reviver, we conducted a within-subject study with 12 PVI. The participants reminisced with two personal photo collections using Memory Reviver and the naïve GPT-4V chatbot (baseline), respectively. Results showed that Memory Reviver enabled the exploration of significantly more scenes ($p < .01$) and helped users recall more memories ($p < .01$) than the baseline. Subjective ratings showed that Reviver outperformed the baseline in facilitating engaging reminiscence, aiding in photo collection understanding, and delivering natural conversational experiences. Moreover, participants reported experiencing strong positive emotions and self-reflection using Memory Reviver. We further discussed directions for personalizing Memory Reviver and summarized implications that could inform future designs of photo reminiscence support and accessible chatbots for PVI.

Our contributions are threefold:

- We present Memory Reviver, a proactive chatbot that leverages two novel features—a *Memory Tree* and a *Proactive Strategy*—to offer guided reminiscence experiences for PVI;
- We contribute an evaluation study that demonstrates how PVI engage with reminiscence using Memory Reviver;
- We distill implications that could guide future designs of photo reminiscence support and accessible chatbots for PVI.

2 RELATED WORK

Our work extends prior research in three areas: (1) photo reminiscence support, (2) image understanding support, and (3) chatbot design for PVI.

2.1 Photo Reminiscence Support for PVI

Reminiscing with photos offers significant psychological benefits, enabling individuals to reflect on the past [17, 36] and envision the future [5, 48]. Prior research has highlighted the desire of PVI to engage in photo reminiscence [29, 37, 82, 84].

To support PVI in reminiscing with individual photos, prior works have explored various tools, including AI-generated image descriptions [37, 84], audio recordings [29], and tactile photography [82]. For instance, Jung et al. [37] investigated the effectiveness of current AI-generated image descriptions for reminiscence and found that these descriptions often focus solely on visual elements (e.g., "*A plate of food on a table*"), making it challenging for PVI to recall associated memories. To address this limitation, Harada et al. [29] proposed linking photos with audio recordings of past moments to assist PVI in reminiscing. Additionally, Yoo et al. [82] suggested using tactile photography to facilitate PVI's engagement with photos. However, these approaches primarily focused on single-photo reminiscence, requiring users to examine each photo individually and manually organize large amounts of information, resulting in a tedious experience.

Recent advances in multi-image-to-text generation techniques have the potential to enable PVI to comprehend multiple images as a whole. These techniques included multi-image summarization [31, 38, 80, 83] and multi-image visual question answering [42, 43, 56]. Recent advances in large multi-modal models (LMM) offer promising opportunities for PVI to comprehend photo collections through natural language input [65, 81]. However, there is a lack of understanding of how these techniques could be leveraged to support their needs. We take an initial step to examine this problem where we invited PVI to reminisce with a photo collection through interacting with a state-of-the-art LMM. Our study uncovered four types of information that PVI seek during reminiscence: a storyline, scene activities, scene details, and new scenes. We further provide practical guidelines on how to distill such information leveraging automated techniques.

2.2 Image Understanding Support for PVI

Enabling image access for PVI has been a consistent research focus in HCI. PVI primarily access images through image descriptions created by manual [7] and automatic methods [67, 77]. Recent studies have investigated how PVI use image descriptions to access online images [75, 84] and how their information needs vary across different scenarios [37, 49, 61, 62]. For example, Stangl et al. [61] found that PVI's preferences for image descriptions vary depending on the image source and the surrounding context.

To fulfill the diverse user needs, several works have explored methods for interactive image exploration, including visual question answering [4, 7, 11, 28] and touch-based image exploration [39, 40, 51, 86]. For example, ImageAssist [32] designed three tools to help PVI explore different regions of interest in an image. However, these works mainly focus on enhancing the exploration of individual images, which is inadequate for complex visual content, such as multiple images [33], multi-page slides [57], and multi-shot videos [66]. Compared to a single image, these complex forms present challenges due to excessive information [33]. To address this challenge, recent research [33, 57, 66] has explored methods to present information associated with complex visual content. For example, GenAssist [33] summarized the similarities and differences among multiple images into an easy-to-compare table format. ShortScribe [66] organized information from short-form videos into hierarchical summaries. However, most existing works focus on presenting information via screen readers, and few works have explored how to convey massive visual information through chat-based interaction. Our work builds on this literature by examining what information in a photo collection facilitates reminiscence (i.e., Memory Tree) and how to present this information via chatbots.

2.3 Chatbot Design for PVI

Chatbots, also known as conversational agents and dialog systems, are software systems designed to simulate human-like conversations through the analysis of natural language data [50]. Due to their conversational nature, chatbots have emerged as promising tools to facilitate accessible interaction for PVI [18, 59]. For example, Pucci et al. demonstrated that chatbots can offer a more natural interaction for web browsing in comparison to screen readers, allowing PVI to directly express their intentions for information retrieval [59]. To inform the design of chatbots for PVI, previous research has proposed general design frameworks [45, 52, 63]. Recent studies have investigated the accessibility issues of voice assistants like Siri [2, 12, 18, 58]. For example, Pradhan et al. [58] found that some PVI struggled to discover the commands supported by a chatbot. However, existing works have primarily focused on designing chatbots to execute specific commands for PVI (e.g., controlling household appliances), leaving a notable gap in the design of chatbots aimed at engaging PVI in reminiscing with photos.

To bridge this gap, we closely involved PVI in our design process. Our findings reveal that the passive "one question, one answer" communication style challenges PVI to engage in the conversation and construct a clear personal story. We thus organize the information needed during reminiscence into a hierarchical structure and offer proactive guidance.

3 FORMATIVE STUDY

To understand the needs and challenges of PVI in reminiscing with a photo collection, we conducted a formative study with eight PVI. The formative study comprised (1) a semi-structured interview to understand their current practices and challenges, and (2) a tech-probe session to investigate their needs and challenges related to reminiscing with a chatbot.

3.1 Methods

Participants. We recruited eight PVI (P1-P8; four males, four females; Table 3 lists demographics) from an online support community. Their ages ranged from 26 to 45 (mean = 32.9, SD = 6.4). Five participants were totally blind and three participants were legally blind with light perception. All participants had previous experience using image accessibility tools (e.g., *Be My AI* [25] and screen readers) and chatbots (e.g., ChatGPT and Siri). They regularly reminisced with photos and consented to share photos for study purposes. All participants were native Mandarin speakers. [1].

[1] In the paper, participants' original speech was translated from Mandarin into English.

Photos. Each participant provided a photo collection related to a specific event, which can be a trip, a gathering, or a performance. Each collection contained 21 to 57 photos (mean = 39.9).

Procedure. The study consisted of two successive phases: a 0.5-hour interview followed by a 1.5-hour tech-probe session, with a break of 5 minutes in between.

Phase 1: Semi-structured Interview. During the interview, we asked participants about their current practices and challenges via the following questions: (1) What's your current practice in reminiscing with photo collections? (2) What are the challenges encountered? (3) What are your expectations for reminiscing with photo collections? To help participants recall previous experiences, we asked about concrete situations, such as "*Can you recall the last time you reminisced with photos?*". When interesting points came up, we followed up with questions to probe additional details.

Phase 2: Tech-probe Session. In this session, participants first received a brief tutorial on a tech-probe chatbot and then used it to reminisce with their photo collections. They conducted free-form conversations with the chatbot. After the session, we conducted an exit interview to understand the participants' needs and challenges in reminiscing with the chatbot. To help them recall their needs and challenges, we read their chat histories with the chatbot aloud to them. The whole study was conducted via a one-on-one Zoom session. Participants were compensated 160 CNY for their time.

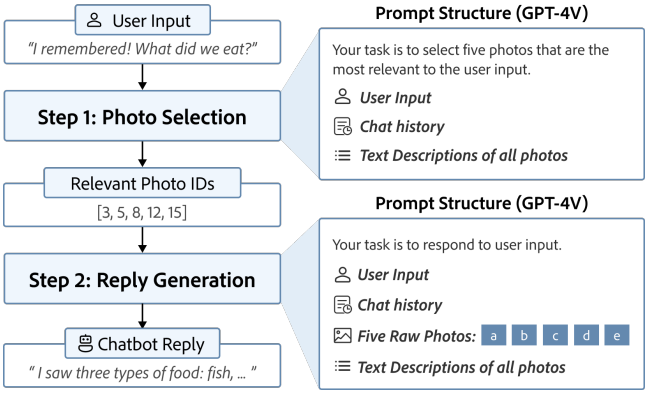

Figure 2: The pipeline of the naïve chatbot.

Tech-probe: a naïve chatbot. We developed a naïve chatbot for the tech-probe session. The chatbot was designed to converse with users based on their utterances and their photo collections. To enable the chatbot to access photo collections, we used GPT-4V to generate natural-language replies based on image inputs [81].

Each time the user provides an input sentence, the system selects the five most relevant photos to generate a reply. Specifically, the system employs a two-step pipeline (see Figure 2). In the first step, it selects five photos by prompting[2] GPT-4V with the text descriptions of all photos. These text descriptions are respectively pre-generated by GPT-4V for each image. Then, in the second step, the system passes the raw data of the five selected photos into GPT-4V to generate a final reply. This pipeline allows the chatbot to access photos in the collection while managing the length of input prompts. We did not directly prompt GPT-4V with the raw images of the entire photo collection due to token limits [81] and the negative impact of long input prompts on performance [13, 72, 76]. Additionally, to enable the chatbot to recognize the user in the photo collection, an extra portrait photo of the user is inputted into GPT-4V.

Users interacted with the naïve chatbot through verbal communication in Mandarin. The chatbot used Amazon Transcribe [3] for speech recognition and Amazon Polly [4] for output speech synthesis. The chatbot operated on the experimenter's desktop and received participants' verbal input directly via Zoom. The average response time[5] of the chatbot was 16.7 seconds (SD=5.3, MAX=31.2), which was reported to be acceptable by the participants.

Analysis. We recorded the study audio, transcribed it, and then qualitatively coded it based on (1) participants' current practice and challenges; (2) their needs and challenges in reminiscing with the chatbot. Additionally, participants' inputs during their conversations with the chatbot were categorized according to their information needs. Two authors conducted the open coding process separately and reached an agreement through discussions. Based on the above data, we report our findings.

3.2 Interview Findings

The semi-structured interview revealed participants' current practices, challenges, and goals during reminiscence.

Current Practices and Challenges. Participants reported saving photo collections for various events, including trips, ceremonies, and stage performances. To make reminiscence accessible, participants adopted common photo-organizing practices. Seven out of eight participants indicated that they intentionally organized photos for each memorable event into separate albums. One participant (P4) mentioned using time frames (e.g., New Year's Day) to filter collections related to specific events.

Despite their efforts in organizing photos, participants faced challenges when reminiscing with photo collections that were "*captured with sighted help*" (P3) and those that were "*taken many years ago*" (P1). The challenges stemmed from their **limited memories** about the photos. Due to their limited memories, participants noted that reminiscing with such a photo collection is "*unachievable without sighted help*" (P3). However, even with sighted help, their needs were often unmet because "*sighted people rarely have the patience to describe the details.*" (P1). Additionally, all participants tried to use AI-image descriptions to reminisce with a photo collection, but they found these descriptions "*too simple to trigger memories*" (P8) and the process too tedious (P2: "*It's tedious to examine each photo one by one, especially when many photos are similar.*"). Consequently, all participants expressed that the current methods were far from effective in supporting reminiscence.

Two common goals. Participants shared two common goals for reminiscence. First, all eight participants expressed a desire to **relive all the scenes** in a photo collection. For instance, P3 expressed,

[2]The prompts are provided in the supplementary materials.

[3]https://aws.amazon.com/pm/transcribe/

[4]https://aws.amazon.com/polly/

[5]This response time refers to the period after the user finishes speaking and before the chatbot starts replying.

"*So many precious scenes were captured during my wedding. I don't want to miss out on any scenes.*". Moreover, five participants hoped to **recall as many memories as possible**. As P2 noted, "*I hope my past memories can be revived to the fullest.*". These two goals informed our task metrics in the evaluation study.

3.3 Tech-probe Findings

The tech-probe session unveiled participants' needs and challenges in reminiscing with a chatbot.

Four Types of Information Needs. When reminiscing with the chatbot, participants commonly reminisced on a scene basis. For example, P7 noted, "*I felt like I was moving through different scenes. I found a scene, explored it to the fullest... and moved on to a new one.*". During the conversations, participants generated a total of 171 inputs. These inputs were categorized into four types[6] based on their information needs: a storyline (13), scene activities (45), scene details (71), and new scenes (28). The needs associated with each information type are as follows:

1. A Storyline: All participants requested the chatbot to generate a storyline encompassing all the scenes in a photo collection. For example, P3 asked the chatbot, "*Could you summarize the whole story in the collection?*" However, participants found that the storyline provided by the chatbot lacked a clear and chronological structure. For instance, P5 mentioned, "*It mixed up many scenes together... and it did not follow a chronological order.*" This finding highlights that **the storyline should be structured as a series of scenes in chronological order (D1)**.

2. Scene Activities: When participants focused on a scene of interest (e.g., "*by the sea*"), they typically started by gathering clues to help them recall past activities: "*recalling what I did is the basis of reminiscence, otherwise photos are just public images to me.*" (P5).

When recalling past activities, participants mainly asked questions regarding "*who, what, when, and where*" (4W) aspects. For example, P3 progressively asked "*Was I alone by the sea? (who)*", "*What was I doing? (what)*", and "*Was it daytime or night? (when)*". Table 1 summarizes the strategies used by participants to uncover the 4W aspects. **One common strategy is to ask about texts in the photos** because "*many texts contain time and location details, such as entrance signs and holiday banners.*" (P1).

However, participants found asking such questions tedious, because "*I didn't know what clues were present in the photos, so I had to try many times.*" (P3). Additionally, participants experienced feelings of doubt and confusion when the reasons for the 4W aspects were not explicitly stated. For example, P2 noted "*I'm confused why the chatbot said it looked like winter. Is it because of our clothing? I need to know the reasons.*". Collectively, findings show the importance of **helping users recall past activities by presenting the "who, what, when, and where" aspects with reasons (D2)**.

3. Scene Details: After recalling past activities, participants proceeded to inquire the chatbot about visual details, such as colors and shapes of the objects existing in the scene. Table 2 summarizes the details inquired by participants, which align with findings in prior works [61, 62]. Their expectations were to enrich their understanding of past experiences: "*All the details reconnected me with the moment: people's poses, their facial expressions... They enrich my memories.*" (P6).

However, they found it hard to fully explore a scene by asking questions: "*Since I can only ask details about objects I already know, I always fear that I have missed out something interesting.*" (P1). While participants tried to ask for a list of details, they found the reply to be overwhelming: "*It is hard to grasp so many details at one time.*" (P4). Collectively, these findings highlight the tension between the need for in-depth exploration and the difficulty in achieving it by asking questions. Thus, it is important for the chatbot to **progressively present scene details to help users fully explore a scene (D3)**.

4. New Scenes: After exploring one scene, participants would find a new scene of interest, so as to "*fully relive all the scenes*" (P5). However, since participants couldn't visually discover new scenes, they relied on the chatbot to provide guidance. For example, P3 asked four times, "*Any other scenes?*". However, participants found that the naïve chatbot either replied "*a random scene*" (P6) or "*a scene that has been discussed*" (P3), leading to confusion about the sequence of scenes and how many were left to explore. This issue stemmed from the memory management of large multi-modal models, a common problem reported in prior works [69, 85]. This finding highlights the importance for the chatbot to **actively suggest new scenes at proper conversation rounds and clearly inform users of the discussion progress (D4).**

Table 1: Strategies used by PVI to recall scene activities.

Aspects	Strategies to recall scene activities from photos
Where	**Landmarks**: e.g., the Eiffel Tower. **Surroundings**: the sea, hills, canteens, museums, etc. **Places in Texts**: entrance signs, holiday banners, etc.
When	**Season**: clothes, tree leaves, etc. **Day or Night**: lighting conditions, etc. **Time in Texts**: holiday banners, screens, etc.
Who	**Visual Appearances**: age, gender, clothes, hair, etc. **Names in Texts**: name tags, name badges, etc.
What	**Human Actions**: e.g., playing musical instruments. **Objects**: food, animals, roller coasters, etc. **Actions in Texts**: menu, conference banners, etc.

Table 2: Scene details asked by PVI.

Aspects	Scene Details asked by PVI
People	Number of people in the photo Gender, age, hair, clothes, facial expression, pose, etc.
Food	Name, color, shape, etc.
Animals	Breed, color, size, etc.
Plants	Species, color, shape, height, etc.
Buildings	Color, shape, style, etc.
Texts	Raw text, position of the text (e.g., on a screen)
Others	Color, shape, etc.

[6]There were 14 inputs classified under "others" which included questions about retouching photos and text-to-image generation.

Two Key Challenges. We identify two key challenges faced by PVI when using the naïve chatbot to reminisce.

The first challenge is the scattering of information throughout the conversation. While participants actively sought the four types of information (D1-D4), the "one question, one answer" communication style led to information being scattered across multiple rounds, making it tedious for participants to recall and organize details. Moreover, the conversation flow also lacks organization. This was evident as the storyline was unstructured and the order of scenes was random and repetitive. This highlights the need for organizing the information into a clear structure.

The second challenge is a lack of proactive guidance from the chatbot. Since participants couldn't visually explore new scenes, the chatbot's lack of proactive guidance further challenged participants to engage in the conversation. As P1 noted, "*Friends would naturally introduce new topics. But with this chatbot, I had to find something new by myself. It didn't feel natural.*" This highlights the importance of the chatbot offering proactive guidance to help users explore a photo collection.

Based on the identified information needs and challenges, we distill two implications for a photo reminiscence chatbot for PVI: (1) **Clear Information Organization**: organize the four types of information into a clear structure. (2) **Proactive Guidance**: proactively lead the conversation and provide information to users at proper conversation rounds. These implications motivated the design of Memory Reviver.

4 MEMORY REVIVER

We present Memory Reviver, a proactive chatbot designed to actively guide users in reliving all scenes within a photo collection. It addresses the two challenges through two novel features: (1) an information architecture of ***Memory Tree***, which organizes the information in a photo collection into a hierarchical structure; and (2) a ***Proactive Strategy***, which actively guides users to explore a photo collection. Together, these features help users gain a clear and thorough understanding of a photo collection during reminiscence.

4.1 Memory Tree

Memory Tree aims to organize information in a photo collection into a structured format. Our formative study reveals four types of information essential for reminiscence: (1) a storyline, (2) scene activities, (3) scene details, and (4) new scenes. Inspired by the organization of autobiographical memory [20], we designed the *Memory Tree* as a three-level tree structure (see Figure 3 (b)).

The Structure of Memory Tree. The *Memory Tree* has three levels: "storyline - scene activity - scene detail". The first level is the **storyline (D1)**, which contains a list of all the scenes arranged chronologically. The second level is the **scene activity (D2)**, which organizes the "who, what, when, and where" aspects into a sentence. These aspects are extracted using the user strategies in Table 1. The third level is the **scene detail (D3)**. Each detail comprises a visual description of objects in the scene (e.g., people, food, animals, etc.), which is extracted according to the guidelines outlined in Table 2. **New scenes (D4)** can be directly retrieved from the storyline. The contents and examples in each level are shown in Figure 3 (c).

Constructing the Memory Tree. Memory Reviver is designed to handle a photo collection about a specific event. After the users upload such a photo collection, the system first arranges the photos in chronological order using the original timestamps in the metadata. It then constructs the *Memory Tree* by segmenting the collection into scenes and distilling information for each scene. To achieve this goal, we leverage the capabilities of GPT-4V [81] in recognizing scenes, activities, landmarks, faces, and texts. We use a three-step pipeline to construct the *Memory Tree* (see Figure 3 (a)):

(1) Scene Segmentation: In our formative study, users perceived a scene to consist of consecutive photos capturing the same activity ("*who, what, when, where*"). Leveraging this insight, we employ GPT-4V to assess the activity similarity between two adjacent photos on a scale of 0 to 1. Empirically, we determine that a segmentation point occurs when the rating is below 0.5. To be noted, this module is not claimed as a contribution in our work and can be substituted with advancements in computer vision [1, 24, 27].

(2) Scene Activity and Detail Extraction: After segmenting the collection into the scenes, we extract information from each scene by inputting the photos into GPT-4V. We prompt[7] the model with the guidelines outlined in Table 1 and Table 2. To ensure the comprehensiveness of the extracted information, all photos from each scene are input together into GPT-4V. Additionally, a portrait photo of the user is inputted for user recognition.

(3) Storyline Generation: After extracting the information for all scenes, we use GPT-4V to generate the storyline. The GPT-4V is prompted with the information of all scenes and the task description (i.e., "*Summarize each scene briefly with a short sentence and then list them in chronological order from the beginning to the end.*")

The pre-extracted information in the *Memory Tree* is then used by the *Proactive Strategy* to guide the conversation.

4.2 Proactive Strategy

The *Proactive Strategy* delivers the information in the *Memory Tree* at proper conversation rounds, forming a natural conversation flow. In achieving this, it employs a state machine [60, 74] to guide the conversation. The state machine is shown in Figure 4 (a).

Proactive Strategy Design. The *Proactive Strategy* starts the conversation with a storyline and then guides users to relive each scene. Within each scene, it introduces the scene activity, progressively presents scene details, and suggests the next scene at proper conversation rounds. The specific conversation flow is as follows:

1. Start the conversation with a storyline: To help users quickly skim through the collection, Memory Reviver starts with a storyline (see Figure 5 (a)). The storyline lists all the scenes chronologically to match users' past experiences **(D1)**. After introducing the storyline, Memory Reviver would suggest starting the exploration of the first scene in the photo collection.

2. Introduce each scene with scene activities: When users enter each scene for the first time, the chatbot will present the scene activity (see Figure 5 (a)). The scene activity encompasses the "who, what, when, and where" aspects **(D2)**. Explanations are provided on how these aspects are determined from photos (e.g., "*It looked*

[7]The prompts are provided in the supplementary materials.

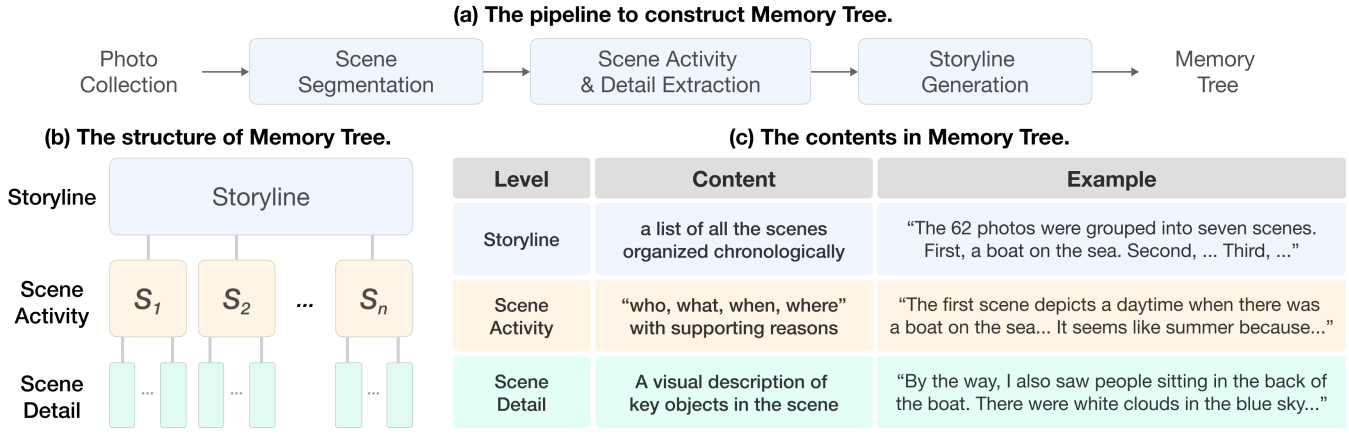

Figure 3: The *Memory Tree* organizes information in a photo collection into a three-level structure.

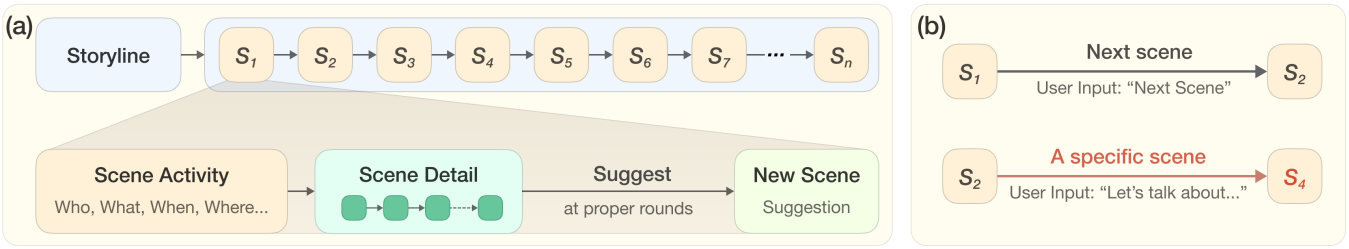

Figure 4: (a) The *Proactive Strategy* starts the conversation with a storyline and then guides users to relive each scene. Within each scene, it introduces the scene activity, progressively presents scene details, and suggests the next scene at proper conversation rounds. (b) Users can freely switch scenes using two natural language commands.

like a canteen because of the 'student canteen' sign at the entrance."). The whole sentence does not exceed 100 characters in length.

3. Present scene details gradually: After the scene introduction, the chatbot will progressively present scene details **(D3)**. In each round of reply, the chatbot will scan through the list of scene details in the *Memory Tree* and present the first detail that has not been discussed in the previous conversation.

4. Suggest new scenes at proper rounds: To help users explore all the scenes, Memory Reviver would actively suggest a new scene at recognized proper conversation rounds **(D4)**. A proper round is defined as when users no longer show interest in the current scene. We adopt the following **criteria to detect this round**: all the scene details in the current scene have been discussed and users have no questions in the last round. Upon detecting this round, Memory Reviver will scan through the storyline and suggest the first scene that has not yet been discussed. This design utlizes the psychological findings that forward recall offers the fastest access to past memories [21]. The scene suggestion is accompanied by a summary of the current scene, which aims to address any remaining questions users may have before switching to the new scene.

5. Allow free scene switch: Apart from actively suggesting the next scene by default, Memory Reviver allows users to flexibly switch among scenes using natural-language commands. Users can employ "*Next scene*" to advance to the next scene or "*Let's talk about ...*" to switch to a specific scene (see Figure 4 (b)).

When all scenes have been discussed, Memory Reviver concludes by providing a summary. This summary serves to help users construct a clear story of their past.

Integrating the Proactive Strategy into Free-form Dialogues. The above design of the *Proactive Strategy* is integrated into free-form conversations, to prevent unnatural interactions — a common challenge in proactive chatbot design [16, 64]. To achieve this goal, Memory Reviver first responds to the user input. Then, it provides proactive guidance by using the *Proactive Strategy* to retrieve information from the *Memory Tree*.

Each time the user speaks an input sentence, Memory Reviver employs a three-stage pipeline to generate a reply (see Figure 5 (b)). The pipeline first (1) identifies a specific scene related to the user input, then (2) generates a raw reply, and finally (3) integrates the proactive guidance into the reply.

(1) Scene Selection: Upon receiving a user input, Memory Reviver determines the current scene using the following rules: (a) If Memory Reviver has suggested a new scene in the last round (including suggesting the first scene at the beginning), it checks whether the user accepts the suggestion by detecting keywords (e.g., "*Okay*", "*Go on*", etc.); (b) If the user input contains "*Next Scene*", Memory Reviver moves directly to the next scene; (c) If the user input contains "*Let's talk about [keyword]*", Memory Reviver

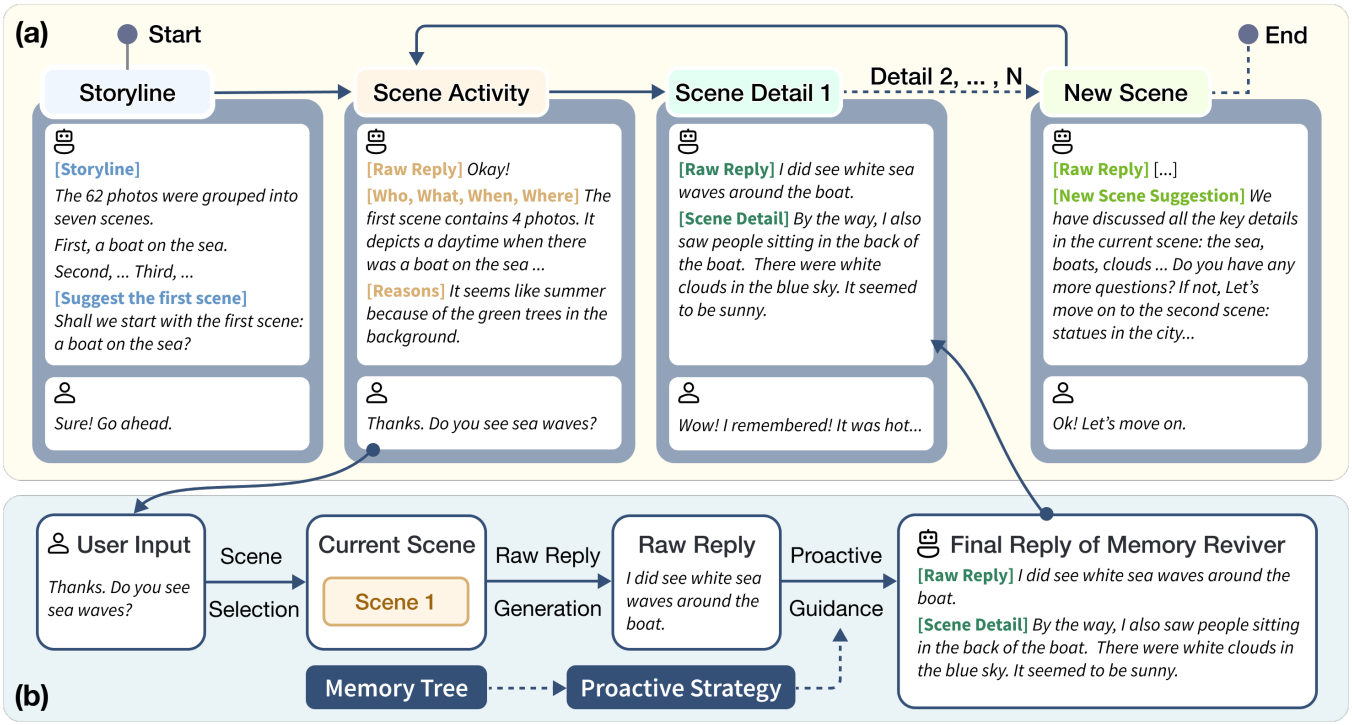

Figure 5: (a) Examples of the multi-round conversations. (b) The pipeline for Memory Reviver to generate a reply in each round.

matches the keyword with the scene details in each scene to identify a specific scene; (d) Otherwise, the scene remains unchanged from the last time (the scene is initially set to scene 1).

The above rules limit Memory Reviver to provide replies on a single-scene basis. We discussed the necessity for addressing cross-scene questions in section 6.2.4 of the evaluation study.

(2) Raw Reply Generation: Memory Reviver uses GPT-4V to generate a raw response. The GPT-4V is prompted[8] with the task description "*Your task is to generate a response to the user input, based on the chat history and the photos.*", along with the user input, the chat history, and the raw photos in the current scene.

(3) Proactive Guidance: Ultimately, Memory Reviver combines proactive guidance with the raw reply. It first retrieves the proactive guidance from the *Memory Tree* using the *Proactive Strategy* outlined in Figure 4 (a). Then, it concatenates the raw reply with the proactive guidance to form a final response. This final reply is subsequently provided to users.

4.3 Implementation

Memory Reviver was implemented as a Python desktop program, with the *Memory Tree* stored in a JSON file and the *Proactive Strategy* implemented as a rule-based state machine [60]. For the AI model, we used gpt-4-vision-preview with a temperature value of 0.8. Users interacted with Memory Reviver through verbal communication in Mandarin. We implemented the voice interaction of Memory Reviver using Amazon Transcribe to recognize users' input speech and Amazon Polly to synthesize output speech.

[8]The prompts are provided in the supplementary materials.

5 TECHNICAL EVALUATION

We evaluated Memory Reviver's technical performance using twelve photo collections provided by PVI. The evaluation focused on two aspects: (1) the scene segmentation performance and (2) the content extraction accuracy.

Materials. We constructed the *Memory Tree* [9] from the twelve photo collections assigned to Memory Reviver in Section 6.1. These collections covered various themes, including trips and ceremonies, with each collection containing between 21 and 68 photos.

Scene Segmentation Performance. To evaluate the scene segmentation, we compared our system's segmentation points with those independently marked by two researchers (Coders M and N). We used the Jaccard index to measure the similarity between each set of segmentation points. The agreement rates were 68% between Coders M and N, 76% between the system and M, and 65% between the system and N. These results show that our system's agreement with human coders was similar to the agreement between human coders. When disagreements occurred, they were mainly due to differing levels of detail (e.g., a single scene by the sea vs. two distinct scenes for walking and eating by the sea).

Content Extraction Accuracy. We annotated inaccuracies by reviewing all photos and statements in the *Memory Tree*. A statement is considered inaccurate if it does not match the photos. One researcher labeled all the data, and a second researcher reviewed the labels to ensure reliability. We calculated the accuracy rates by dividing the number of correct statements (i.e., statements without inaccuracies) by the total number of statements. The accuracy rates

[9]The Memory Tree data is provided in the supplementary materials.

were 96.4% for storylines, 95.2% for scene activities, and 90.8% for scene details. The major error cases were the false identification of objects (14 times), texts (9 times), and human genders (5 times). These errors occurred due to the hallucination of GPT-4V [81].

6 USER EVALUATION

We conducted a user study with 12 PVI to evaluate the effectiveness of Memory Reviver. Our evaluation focuses on three aspects: (1) reminiscence experience, (2) understanding of a photo collection, and (3) conversational experience.

6.1 Methods

In a within-subject study, participants used Memory Reviver and a baseline chatbot to reminisce with personal photo collections.

Participants. We recruited 12 PVI (six males, six females) who hoped to reminisce with photo collections (P7-P18, Table 3). These participants were recruited from an online support community. Their ages ranged from 26 to 52 (mean = 33.9, SD = 7.4). All participants were either legally blind or totally blind, and they utilized image accessibility tools such as *Be My AI* [25] and screen readers. Participants had experiences with chatbots such as ChatGPT, Gemini, and Siri. P7 and P8 took part in the formative study.

Photo Collections. All participants agreed to share personal photos for study purposes. To ensure study control, each participant provided two collections meeting the following criteria: (1) the two collections documented two distinct events; (2) both events shared the same theme (e.g., family trips) and were equally significant to the participant; (3) the difference in the number of photos between the two collections was less than five; and (4) participants had limited memory of the photos and the events.

To ensure generalizability, the collections provided by different participants covered a wide range of cases: (1) their themes included trips, gatherings, ceremonies, and conferences, and (2) the number of photos in each collection ranged from 21 to 68 (mean = 37.2).

Baseline. To assess the effect of integrating user insights into Memory Reviver, we used the naïve chatbot (used in the formative study, Figure 2) as our baseline.

Task and Procedure. We first gathered the participants' demographics and asked about their current practices of reminiscing with photos. Following this, the participants received a 10-minute tutorial on both Memory Reviver and the baseline chatbot using a fixed set of public images. They conducted free-form conversations with each chatbot. After the tutorial, the participants took part in two reminiscence sessions.

The **task** in each session was to use a chatbot to reminisce with a photo collection, with the two goals identified in the formative study (section 3.2): (1) exploring all the scenes and (2) recalling associated memories. The order of the chatbots was counterbalanced and the photo collections were randomly assigned to participants. Each session was conducted as follows: (1) First, participants verbally composed a pre-trial **memory narrative**, prompted by the folder name of the photo collection. They were instructed to recall all memories associated with the photos. To ensure study control, no interaction with or feedback from the experimenter was provided during this narrative. (2) Next, the participants engaged in free-form conversations with the assigned chatbot. The conversation continued until the participants had no more replies. (3) After the conversation, participants composed a post-trial memory narrative using the same instructions as the pre-trial narrative. They then received a post-trial survey, which included the ratings in Figure 7. All ratings were on a 7-point Likert scale.

There was a 5-minute break between each session. At the end of the study, a semi-structured interview was conducted to collect participants' feedback on three aspects: (1) reminiscence experience; (2) photo collection understanding; and (3) conversational experience. The study lasted 2.5 hours and was conducted via Zoom in a one-on-one session. Both chatbots operated on the experimenter's desktop and received participants' verbal input directly via Zoom. The participants were compensated 200 CNY for their time.

Metrics. Informed by the two goals identified in the formative study (section 3.2), We adopted two task metrics:

(1) **Memory Ratio** evaluates the effectiveness of the chatbot in helping users recall memories. Specifically, *Memory Ratio = (Word count of post-trial narrative) / (Word count of pre-trial narrative).* The use of word count to measure memory recall was commonly adopted in prior studies [5, 55].

(2) **Scene Coverage** evaluates whether the chatbot fulfilled the users' goal of exploring all the scenes in a photo collection. Specifically, *Scene Coverage = (Number of scenes discussed) / (Number of all scenes).* All photo collections were segmented into scenes using the method in Figure 4. For Memory Reviver, a scene was considered discussed if the scene was selected for reply generation. For the baseline chatbot, a scene was considered discussed if at least one photo within the scene was selected for reply generation.

Besides task metrics, we used subjective ratings to measure user experience in three aspects, as shown in Figure 7.

Analysis. We recorded the study audio, participants' memory narratives, their chat history with the chatbots, and their subjective ratings. The memory narratives were transcribed verbatim as they were spoken to tally their word count. The interview audio was transcribed and categorized according to the three aspects of the interviews. Based on the data, we report our findings.

6.2 Results

We report results in three aspects: (1) reminiscence experience, (2) understanding a photo collection, and (3) conversational experience.

6.2.1 ***Reminiscence Experience.*** Memory Reviver was shown to be effective in facilitating memory recall and supporting enjoyable reminiscence. Moreover, we found that Memory Reviver even elicited strong positive emotions and in-depth self-reflection from some participants. The detailed results are as follows:

First, Memory Reviver was more effective in aiding in memory recall than the baseline. Memory Reviver achieved a memory ratio of 5.43, indicating that participants narrated an average of 5.43 times more words in their post-trial memory narratives than in their pre-trial narratives. This result significantly surpassed the baseline's memory ratio of 1.79 ($\mu = 5.43, \sigma = 2.61$ vs. $\mu = 1.79, \sigma = 1.39$; $t_{11} = 6.47, p < .01$). Participants rated Memory

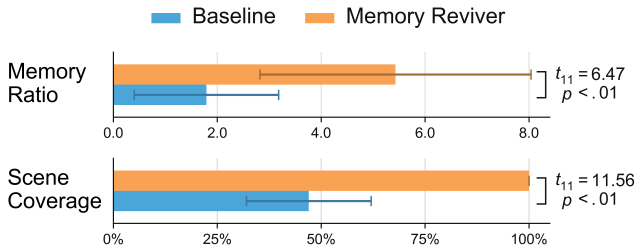

Figure 6: Task Performances in the evaluation study. Paired t-test was used for significance analysis.

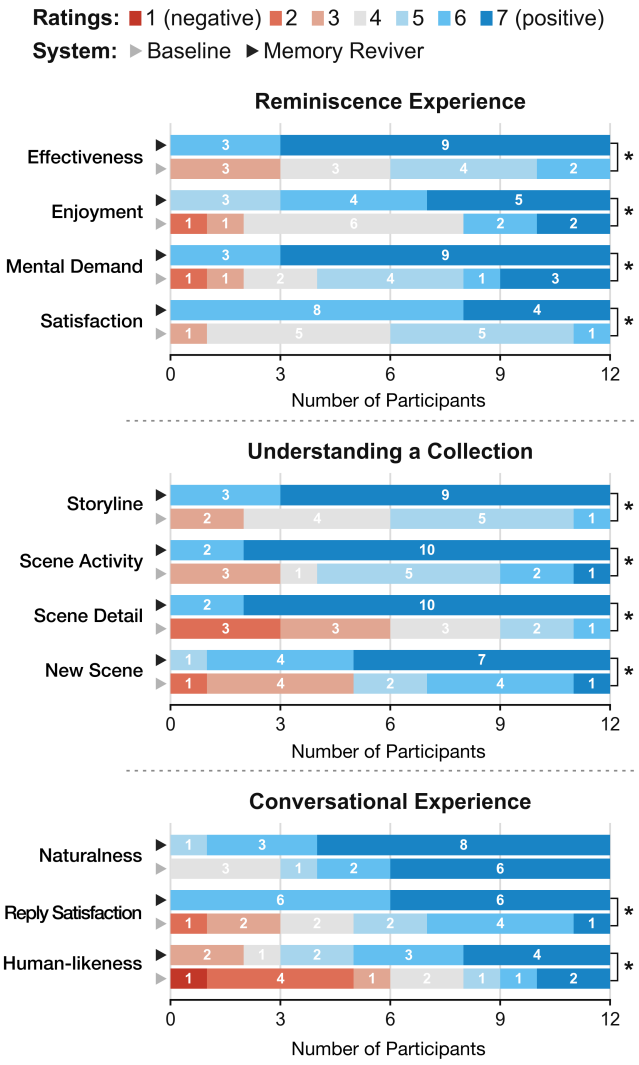

Figure 7: Distributions of the ratings for Memory Reviver and the baseline (1=negative, 7=positive). The asterisks indicate the statistical significance as a result of the Wilcoxon signed-rank test (* denotes p < 0.01).

Reviver as more effective in helping them fully recall past memories ($\mu = 6.75, \sigma = 0.45$ vs. $\mu = 4.42, \sigma = 1.08$; $Z = -3.08, p < .01$). They attributed their effective memory recall not only to the well-organized information presented by Memory Reviver but also to the low-mental-demand experience of being guided to relive past moments. As P12 noted, "*The guidance was especially useful in evoking long-lost memories... I felt fully immersed in my past.*".

Second, Memory Reviver delivered a more enjoyable experience than the baseline. Participants rated Memory Reviver higher in enjoyment ($\mu = 6.17, \sigma = 0.83$ vs. $\mu = 4.58, \sigma = 1.56$; $Z = -2.70, p < .01$), lower in mental demand ($\mu = 6.75, \sigma = 0.45$ vs. $\mu = 5.00, \sigma = 1.60$; $Z = -2.68, p < .01$), and with higher overall satisfaction ($\mu = 6.33, \sigma = 0.49$ vs. $\mu = 4.50, \sigma = 0.80$; $Z = -3.17, p < .01$). During the exit interview, all participants expressed a preference for using Memory Reviver over the baseline for future reminiscence, because it is more engaging (mentioned 12 times) and less mentally demanding (9 times). Interestingly, Participants likened their experiences with Memory Reviver to entertaining activities such as "*watching a movie*" (P9, P17), "*taking a tour*" (P15), or "*playing a side-scrolling game*" (P11). In contrast, they associated the baseline with tasks such as "*retrieving information*" (P13), "*taking a test*" (P12), or "*finding ways in a maze*" (P11). Participants ascribed this enjoyable and low-mental-effort experience to both the easy understanding of photo collections (mentioned 12 times) and the natural conversational experiences (9 times).

Surprisingly, Memory Reviver elicited strong positive emotions and in-depth self-reflection for several participants. Five participants reported experiencing strong positive emotions after reminiscing with Memory Reviver. For example, P10 stated, "*I've never imagined my memories can be revived so vividly. It made my eyes well up with tears of joy.*" Seven participants reported reflecting on life experiences beyond photos, such as career planning and family relationships. Participants attributed their in-depth reminiscence with Memory Reviver to the low mental demand required to understand the photos, allowing them to "*fully engage in reminiscence*" (P11). The implications of these findings are further discussed in section 7.3.

6.2.2 ***Understanding a Photo Collection.*** We examined whether Memory Reviver could support the users' goal of exploring all the scenes in a photo collection. We also evaluated the effectiveness of Memory Reviver in supporting the four information needs identified in the formative study. The results are as follows:

All participants explored 100% of scenes when using Memory Reviver, which is significantly higher than the 47% scene coverage achieved with the baseline ($\mu = 100\%, \sigma = 0\%$ vs. $\mu = 47\%, \sigma = 15\%$; $t_{11} = 11.56, p < .01$). This indicates that Memory Reviver successfully enabled participants to explore all scenes in a collection. Participants ascribed this 100% scene coverage of Memory Reviver to the proactive guidance: "*It listed all the scenes in the beginning, and actively guided me to explore each one. With such guidance, I felt like I was taking a tour and couldn't wait to see them all.*" (P15). In contrast, when using the baseline, participants all ended the chat after several unsuccessful attempts to discover new scenes, as "*It often repeated what we've already discussed.*" (P12).

Participants rated Memory Reviver as significantly higher in fulfilling their information needs, including a clear storyline

($\mu = 6.75, \sigma = 0.45$ vs. $\mu = 4.42, \sigma = 0.90$; $Z = -3.11, p < .01$), easy recall of past activities ($\mu = 6.83, \sigma = 0.39$ vs. $\mu = 4.75, \sigma = 1.29$; $Z = -2.88, p < .01$), easy exploration of scene details ($\mu = 6.93, \sigma = 0.39$ vs. $\mu = 3.58, \sigma = 1.31$; $Z = -3.07, p < .01$), and easy discovery of new scenes ($\mu = 6.50, \sigma = 0.67$ vs. $\mu = 4.58, \sigma = 1.68$; $Z = -2.97, p < .01$). Their feedback is as follows:

1. A Clear Storyline: All participants praised the storyline for helping them recall past memories. For example, P9 noted "*The storyline was so clear that it felt like watching a movie of my past.*". Participants also highlighted the importance of organizing the scenes in chronological order: "*If it had ignored the timeline and grouped the photos by people or objects, I couldn't have recalled my past experiences so clearly.*" (P14). Conversely, a lack of a well-organized storyline in the baseline resulted in a higher mental demand: "*It's like taking a test. I have to manually piece together all those memory fragments in my mind, which is quite hard.*" (P11).

2. Scene Activities that Aided in Memory Recall: Participants noted that the scene activities (i.e., "*who, what, when, and where*") effectively helped them recall past activities: "*Instead of just describing the contents and requiring me to guess what I did, it directly predicted my past activities. These descriptions instantly brought me back to the moment.*" (P7). Participants particularly praised Memory Reviver's ability to provide supporting reasons: "*It told me the place might be either a cafe or a canteen according to the coffee cup on the table. Not even a friend has ever given such precise predictions.*" (P13). Some participants even stated that they learned new strategies to understand their photos: "*I was pretty surprised when it told me 'the name of the place might be ... according to the text on the door.' It taught me new ways to find clues in my photos.*" (P9). In contrast, when using the baseline, participants typically asked several consecutive questions to recall their past activities, and as a result, they noted, "*It was time-consuming because I needed to figure out how the photos related to my past memories all by myself.*" (P15).

3. Scene Details that Enriched Memories: Participants expressed that the active presentation of scene details gave them delightful surprises and enriched their memories. As P15 noted, "*I had the impression of taking photos there, but I had no idea so many beautiful buildings and plants were captured behind me. It truly enriched my memories.*". The active presentation of new details made some participants feel "*warm*" (P10, P15). Besides enriching their memories, they also mentioned that the progressive presentation of details reduced their cognitive load: "*While some tools may offer highly detailed descriptions, I often find them overwhelming. This chatbot (Memory Reviver) solved this issue for me.*" (P13). Conversely, with the baseline, participants felt like "*retrieving information*". As P8 noted, "*It only answered my questions, but how could I continue asking questions if I'm not learning anything new?*".

4. New Scene Suggestions that Reduced Mental Efforts: Participants stated that the scene suggestions reduced their mental efforts: "*I didn't need to keep track of the discussion progress myself. I simply let it take the lead so that I could fully immerse myself in my memories.*" (P18). Additionally, participants also noted that the summary of the current scene before suggesting the next scene was especially helpful (e.g., Memory Reviver: "*We have talked about all the key contents in the current scene: ... Do you want to proceed to the next scene?*"), because "*such a summary made me feel confident that I didn't miss out anything interesting.*" (P7).

6.2.3 ***Conversational Experience.*** Participants generated 338 inputs when using Memory Reviver and 167 inputs with the baseline. The average response time of the chatbots was 14.6 seconds for Memory Reviver (SD=4.8, MAX=27.9) and 13.8 seconds for the baseline (SD=4.5, MAX=31.6). Regarding the conversation flow, eight participants explored all the scenes from the first to the last, and the other four participants actively switched to a specific scene (P7: 2 times; P12, P14, P17: 1 time) because "*a new scene just came into my mind*" (P7, P14, P17) or "*I hoped to ask more about a previous scene*" (P7, P12). We assessed the conversational experience in three aspects: (1) whether the proactive guidance of Memory Reviver would lead to unnatural conversation flow, a common concern in proactive chatbot design [16, 64]; (2) users' satisfaction with the chatbot's replies; and (3) the chatbot's ability to simulate a human-like experience. Based on users' ratings and feedback, we have the following findings:

First, Memory Reviver delivers a natural conversation flow by balancing proactive guidance with user freedom. Participants rated Memory Reviver higher on average than the baseline in terms of providing a natural conversation flow ($\mu = 6.58, \sigma = 0.67$ vs. $\mu = 5.92, \sigma = 1.31$; $Z = -1.63, p = 0.10$), although no statistically significant difference was identified. They attributed this natural flow to the sense of control they had over the conversation with Memory Reviver: "*All of its guidance made me more aware of the photos. Even if some suggestions didn't interest me, I could simply ask about what I wanted to know. I was in control of the conversation.*" (P16). Similarly, P7 noted, "*It had a default storyline, but I had total freedom to switch to another scene.*". In contrast, participants expressed lacking control when using the baseline. As P7 noted, "*Although it seemed that we can freely communicate with the chatbot (the baseline), I often felt lost because I didn't know where our conversation was headed.*". Collectively, findings suggest that Memory Reviver delivered a natural conversation flow by achieving a balance between providing proactive guidance and allowing users the freedom to lead the conversation.

Second, Memory Reviver provided satisfactory replies by actively introducing new information. Memory Reviver significantly outperformed the baseline regarding reply satisfaction ($\mu = 6.50, \sigma = 0.52$ vs. $\mu = 4.75, \sigma = 1.54$; $Z = -2.54, p < .01$). Participants consistently attributed Memory Reviver's higher reply satisfaction to its active presentation of new information. For example, P10 noted, "*It (Memory Reviver) not only responded to my words but also mentioned something new in the photos. This kept our conversation going.*" In contrast, participants found the baseline to provide limited information after several rounds of discussion: "*It often got stuck with a similar topic repeatedly. It never said 'Oh, I had something new for you.'. This made me feel bored.*" This finding highlights the importance of introducing new information to deliver a satisfying conversational experience.

Third, Memory Reviver delivered a human-like experience. Participants rated Memory Reviver as significantly more human-like than the baseline ($\mu = 5.50, \sigma = 1.51$ vs. $\mu = 3.75, \sigma = 2.09$; $Z = -2.70, p < .01$) mainly due to its proactive guidance. As P15 noted, "*It's like a friend who is willing to guide me through the photos... It also learned from my past experiences. I felt warm.*". During their chat with Memory Reviver, nine participants spontaneously shared their past experiences with Memory Reviver, because "*It*

elicited my hope to share my past experiences." (P18). For example, P11 shared his anecdotes with Memory Reviver: "*You know what? We got lost during that trip ...*". Additionally, eleven participants (all except P17) highlighted that their experiences with Memory Reviver surpassed any previous reminiscence with sighted friends. They credited this to the sufficient details provided by the chatbot (P10: "*No sighted friends have the patience to provide so many details.*") and its dedicated services (P15: "*Chatbots always prioritize our needs, but friends may not.*"). However, participants noted that Memory Reviver's tone and response time (mean = 14.6 seconds) still presented barriers to achieving a human-like experience.

6.2.4 ***Concerns and Improvements.*** We report the concerns and suggested improvements identified during the evaluation study in the following three aspects:

(1) Incorrect Information: Both chatbots provided incorrect information (Memory Reviver: 19 out of 338 replies; the baseline: 12 out of 167 replies), including false identification of users, visual details, and hallucinations. These errors were counted in real time by the experimenter and confirmed after the trial. Most of the errors were unnoticed during the conversation because "*I had little memory of the photos, so I accepted everything it said.*" (P15). A few errors were noticed due to mismatches with their memories (P14: "*I never had a wrist-watch!*") or inconsistencies across consecutive rounds of reply (P14: "*It gave me two different texts.*"). Moreover, we found that **slightly different descriptions of colors and shapes could lead to confusion**. For example, P11 noted "*The chatbot mentioned me wearing a pink dress, then a light pink one. Did I change clothes?*"). This highlights the importance of using consistent descriptions for identical objects.

After the trial, the experimenter corrected any misinformation for the participants. Eleven participants expressed tolerance of occasional errors because "*The joy of reminiscence outweighed the errors.*" (P13). One participant (P17) cautioned against any errors because "*I only keep photos of memorable events. If the chatbot cannot be accurate, I'd prefer to reminisce with my partner.*".

(2) Subjective Information: All participants showed high acceptance of using AI-generated storyline and activity descriptions for reminiscence. While such information tends to be subjective and may not perfectly align with users' past experiences, participants expressed few concerns. They highlighted two main reasons: First, they believed that no other individuals could perfectly describe their past experiences: "*Even different friends describe my photo collections differently.*" (P11). Second, they mentioned that the supporting reasons effectively alleviated their concerns about subjectivity (e.g., "*You seemed to be in a zoo because of the 'no feeding animals' sign.*"). This finding indicates that combining factual contents with subjective descriptions is a potential solution to reduce PVI's concerns about subjectivity, a common issue mentioned in prior works [37, 61, 62].

(3) Future Improvements of Memory Reviver: Participants provided suggestions for improving Memory Reviver. The most common suggestion (mentioned 7 times) was for Memory Reviver to have long-term memories across different reminiscence sessions. For example, P13 noted, "*I want it to learn that this dog is my guide dog not only for this conversation but for all future ones.*". Five participants noted that Memory Reviver sometimes suggested new information that had already been mentioned earlier in the conversation. To address this issue, future improvements should focus on filtering out redundant information that has already been discussed before adding the proactive guidance. Four participants suggested that Memory Reviver should be able to respond to questions related to multiple scenes (e.g., "*Is the dog on the beach the same as the one on the grass?*"). Additionally, three participants suggested customizing the role of the chatbot or its narrative styles, which we discuss in section 7.1.

7 DISCUSSION

We reflect on findings from the design and evaluation of Memory Reviver and discuss the implications for future design.

7.1 Personalization of Memory Reviver

Based on the evaluation study, we identify several future directions to tailor Memory Reviver to different users' preferences.

Personalize responses using interaction histories. While PVI could ask diverse visual questions [37, 62], we noticed that the same participant in the user study tended to ask similar questions when viewing photos in a collection. For instance, P16 frequently inquired about facial expressions (14 out of 21 questions), while P13 often asked about body postures (8 out of 17). Our preliminary observation suggests that this tendency might be related to both the photo themes (e.g., a family trip) and their personal preferences. This indicates an opportunity to learn from user preferences and adjust subsequent responses using online learning methods [30].

Prioritize information based on user contexts. In our user study, participants indicated that their reminiscence interests might depend on their current contexts, such as locations (P11: "*When revisiting a country, I like to review photos taken there previously.*"). This suggests that Memory Reviver could prioritize content based on user contexts like location [48] and time [55].

Reduce the frequency of suggestions for older adults. Currently, Memory Reviver provides proactive guidance in every reply. While eleven participants found this frequency appropriate, P9 (aged 52) occasionally experienced an interruption of thought due to the new information in the suggestions. This could be linked to age-related cognitive declines [23]. Given that older adults often have a strong desire for reminiscence [55], it's essential to customize the frequency of suggestions to their needs.

Tailor the role of the chatbot. Participants in our evaluation study had differing views on the role of the chatbot. Four participants envisioned the chatbot as a friend with a consistent personality: "*I hope it could chat with me about its own experiences.*" (P10). Conversely, eight participants preferred the chatbot to act as an information provider, as they viewed reminiscence as a personal experience: "*It serves to provide information and I'll reminisce by myself.*" (P17). This indicates the role and personality of the chatbot [54, 68] could be further customized.

Customize description styles. Participants held diverse preferences regarding the description styles of the chatbot. For example, P15 expected the chatbot to provide subjective descriptions such as "*You seemed really happy.*", while P17 preferred more objective descriptions. This indicates that the description styles could be further customized by leveraging insights from prior works [49].

7.2 Privacy and Ethical Considerations

Our research involved using personal photos to elicit personal memories. To address the associated privacy concerns, we carefully took measures in both user studies. First, we informed participants of the privacy risks associated with submitting photos to a third-party service and obtained their consent. Second, we removed all metadata from the photos and ensured that data used in API calls would not be used for training purposes. Third, we submitted data deletion requests after the studies. We acknowledge that some privacy risks, such as data leaks by third-party services, remain.

To eliminate the privacy risks, future work should focus on using local visual language models on mobile devices [19, 34] to enhance data privacy. Moreover, reminiscence may trigger memories of negative experiences [35]. To minimize such negative effects, future systems should enable users to filter out certain content before starting reminiscence sessions.

7.3 Design Implications

Photo Reminiscence Support. During the user evaluation, participants mentioned the unique benefits of reminiscing with a chatbot: offering patient and dedicated services, reducing social concerns, and promoting self-reflection. Notably, seven participants spontaneously engaged in self-reflection after using Memory Reviver. For example, P7 contemplated career planning, P8 decided to reconnect with long-lost friends, P13 resolved to spend more time with her parents, and P16 expressed the intent to resume playing the piano. Their self-reflection indicated the depth of reminiscence [35]. Participants noted that such depth was rare in their previous reminiscence with photos. We identified three contributing factors to this in-depth experience. (1) The **ease of understanding a photo collection** allowed them to engage in self-reflection. As P7 noted, "*I naturally reflected on my life choices because understanding many photos was no longer tedious.*". (2) The **rich details** helped them deeply connect with and reflect on past experiences: "*After hearing how splendid the music hall was, I wanted to perform there again.*" (P16). (3) The **independent reminiscence** facilitated their self-reflection. As P8 noted, "*When I reminisce with friends, it's more about chit-chat than reflection.*".

Collectively, our study reveals the significant potential of photos in facilitating self-reflection for PVI. This finding suggests that future photo reminiscence support should not only focus on helping PVI understand photo contents [37, 82, 84], but also aim to engage them in self-reflection [6]. By alleviating the cognitive load of understanding photos, providing rich details, and supporting independent reminiscence, accessibility tools can help PVI deeply engage with photos and derive psychological benefits.

Chatbot Design for PVI. Our study revealed that PVI aimed to fully explore a photo collection during their conversations with a chatbot. This objective differs from chat-based information retrieval [41] or question answering [4], as it requires users to actively organize and synthesize large amounts of information. However, achieving this task poses challenges due to the limited memory storage capacity of auditory channels [15]. To address this challenge, our approach is to (1) distill PVI's information needs by observing their natural conversations with a chatbot, (2) organize the essential information into a clear structure, and (3) gradually present information to users [32, 49, 57]. User evaluation demonstrated the effectiveness of our design.

Similar to this task, PVI also need to comprehend large amounts of information for tasks such as reading [44] and web browsing [59]. Drawing upon our findings and established guidelines [16, 53], we derive implications for designing accessible chatbots aimed at supporting the comprehension of large amounts of information as follows: (1) Identify users' information needs and organize information into a clear structure, (2) Offer an overview to help users quickly skim the information, (3) Present information progressively to alleviate cognitive load, and (4) Summarize information regularly to mitigate users' fear of missing out (e.g., Memory Reviver provides a summary of the current scene before suggesting a new scene.). These implications extend prior literature on accessible chatbots for PVI [2, 12, 18, 58, 59], offering practical insights for designing chatbots tailored to the needs of PVI when consuming large amounts of information.

Additionally, when designing assistive chatbots for real-time scenarios such as outdoor navigation [46, 78, 79] and shopping [10], it is vital to avoid overwhelming users with excessive information [10]. Our strategy of presenting information in an "overview first, details on demand" manner could help reduce the associated cognitive load in these scenarios.

7.4 Limitations and Future Work

We summarize the limitations and future work in two aspects: (1) the scope of Memory Reviver, and (2) limitations of user studies.

Scope of Memory Reviver. Memory Reviver is designed to help PVI reminisce with a photo collection using a chatbot. Its scope is bounded by the use of conversational user interfaces, the types of photo collections it can process, the reminiscence needs it caters to, and the performance of its underlying models.

First, Memory Reviver leverages natural conversation with a chatbot to facilitate reminiscence. Future work should explore additional design possibilities, such as utilizing multi-modal input (e.g., touch-based image exploration [33, 51]) and multi-sensory feedback (e.g., auditory and tactile feedback [29, 70, 82]), to further enhance the reminiscence experience.

Second, Memory Reviver is tailored for photo collections associated with specific events like trips, family gatherings, ceremonies, and stage performances (see Table 3). Such collections typically comprise dozens to hundreds of photos that can be segmented into multiple scenes. To extend Memory Reviver's ability in processing larger and more diverse photo libraries [17], future work should explore methods to automatically select photos for reminiscence [34, 48] and use metadata to filter out outliers like screenshots. Additionally, we noticed that users' inquiries may be influenced by the themes of photo collections. Therefore, fine-tuning proactive suggestions based on the specific themes of photo collections could be a potential direction for future work.

Third, Memory Reviver is designed to accommodate the reminiscence needs of fully exploring a photo collection. Currently, it restricts exploration on a scene basis and does not support cross-scene queries, which we view as promising future work. Moreover, users may have other needs when interacting with a photo collection, such as retrieving a specific photo [29, 37]. While the specific

solution may not be directly applicable, our methodology of delivering proactive guidance according to user needs can be leveraged.

Fourth, Memory Reviver inherits the limitations of its underlying models. The use of GPT-4V [81] occasionally introduced image recognition errors and hallucinations, which could be mitigated by incorporating specialized models (e.g., precise text recognition [9]) and model advancements (e.g., alleviating hallucinations [8]).

Fifth, Memory Reviver currently employs rule-based methods to select proactive suggestions from pre-extracted information (i.e., the *Memory Tree*). Future works include improving the suggestion methods and customizing pre-extracted information according to diverse user preferences.

Limitations of User Studies. First, participants in our evaluation study reminisced with their personal photo collections. While we attempted to control for variations between collections, differences could not be completely eliminated. Second, the majority of participants in our user studies were young adults, with only one older participant (P9, aged 52). As such, the effectiveness of Memory Reviver for older adults requires further investigation. Third, we did not evaluate the long-term effects of Memory Reviver in supporting daily reminiscence, which we see as potential future work.

8 CONCLUSION

Memory Reviver is a proactive chatbot designed to assist people with visual impairment in reminiscing with a photo collection. It addresses two key challenges hindering effective reminiscence with a chatbot: the scattering of information and a lack of proactive guidance. Through user evaluation, we demonstrated the effectiveness of Memory Reviver in facilitating engaging reminiscence, enhancing understanding of photo collections, and delivering natural conversations. We identified future directions to personalize Memory Reviver according to diverse user needs and distilled implications for photo reminiscence support and chatbot design for PVI. We hope this work will offer useful insights into the design of accessible chatbots and inspire researchers to develop tools that facilitate enjoyable and in-depth reminiscence.

ACKNOWLEDGMENTS

The authors would like to thank the participants for their support during the studies. Additionally, we thank the reviewers for their constructive feedback. We thank Prof. Chun Yu from Tsinghua University and Prof. Guanhong Liu from Tongji University for their methodological guidance. We thank Qiyue Cai from Arizona State University for introducing the structure of autobiographical memory, which inspired the design of *Memory Tree*. This work is partially supported by the Research Grants Council of the Hong Kong Special Administrative Region under General Research Fund (GRF) with Grant No. 16214623.

REFERENCES

[1] Sathyanarayanan N Aakur and Sudeep Sarkar. 2019. A perceptual prediction framework for self supervised event segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 1197–1206.

[2] Ali Abdolrahmani, Ravi Kuber, and Stacy M. Branham. 2018. "Siri Talks at You": An Empirical Investigation of Voice-Activated Personal Assistant (VAPA) Usage by Individuals Who Are Blind. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility* (Galway Ireland). ACM, 249–258. https://doi.org/10.1145/3234695.3236344

[3] Rebecca S Allen. 2009. The legacy project intervention to enhance meaningful family interactions: Case examples. *Clinical Gerontologist* 32, 2 (2009), 164–176.

[4] Stanislaw Antol, Aishwarya Agrawal, Jiasen Lu, Margaret Mitchell, Dhruv Batra, C Lawrence Zitnick, and Devi Parikh. 2015. Vqa: Visual question answering. In *Proceedings of the IEEE international conference on computer vision*. 2425–2433.

[5] Benett Axtell, Raheleh Saryazdi, and Cosmin Munteanu. 2022. Design is worth a thousand words: The effect of digital interaction design on picture-prompted reminiscence. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–12.

[6] Marit Bentvelzen, Jasmin Niess, and Paweł W Woźniak. 2021. The technology-mediated reflection model: Barriers and assistance in data-driven reflection. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–12.

[7] Jeffrey P Bigham, Chandrika Jayant, Hanjie Ji, Greg Little, Andrew Miller, Robert C Miller, Robin Miller, Aubrey Tatarowicz, Brandyn White, Samual White, et al. 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.

[8] Ali Furkan Biten, Lluís Gómez, and Dimosthenis Karatzas. 2022. Let there be a clock on the beach: Reducing object hallucination in image captioning. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 1381–1390.

[9] Ali Furkan Biten, Ron Litman, Yusheng Xie, Srikar Appalaraju, and R Manmatha. 2022. Latr: Layout-aware transformer for scene-text vqa. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 16548–16558.

[10] Roger Boldu, Denys JC Matthies, Haimo Zhang, and Suranga Nanayakkara. 2020. AiSee: an assistive wearable device to support visually impaired grocery shoppers. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 4 (2020), 1–25.

[11] Erin Brady, Meredith Ringel Morris, Yu Zhong, Samuel White, and Jeffrey P Bigham. 2013. Visual challenges in the everyday lives of blind people. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 2117–2126.

[12] Stacy M. Branham and Antony Rishin Mukkath Roy. 2019. Reading Between the Guidelines: How Commercial Voice Assistant Guidelines Hinder Accessibility for Blind Users. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility* (Pittsburgh PA USA). ACM, 446–458. https://doi.org/10.1145/3308561.3353797

[13] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems* 33 (2020), 1877–1901.

[14] Fred B Bryant, Colette M Smart, and Scott P King. 2005. Using the past to enhance the present: Boosting happiness through positive reminiscence. *Journal of Happiness Studies* 6 (2005), 227–260.

[15] Stuart K Card. 2018. *The psychology of human-computer interaction*. Crc Press.

[16] Ana Paula Chaves and Marco Aurelio Gerosa. 2021. How should my chatbot interact? A survey on social characteristics in human–chatbot interaction design. *International Journal of Human–Computer Interaction* 37, 8 (2021), 729–758.

[17] Amy Yo Sue Chen, William Odom, Carman Neustaedter, Ce Zhong, and Henry Lin. 2023. Exploring Memory-Oriented Interactions with Digital Photos In and Across Time: A Field Study of Chronoscope. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg Germany). ACM, 1–20. https://doi.org/10.1145/3544548.3581012

[18] Dasom Choi, Daehyun Kwak, Minji Cho, and Sangsu Lee. 2020. " Nobody speaks that fast!" An empirical study of speech rate in conversational agents for people with vision impairments. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.

[19] Xiangxiang Chu, Limeng Qiao, Xinyu Zhang, Shuang Xu, Fei Wei, Yang Yang, Xiaofei Sun, Yiming Hu, Xinyang Lin, Bo Zhang, et al. 2024. Mobilevlm v2: Faster and stronger baseline for vision language model. *arXiv preprint arXiv:2402.03766* (2024).

[20] Martin A Conway and Debra A Bekerian. 1987. Organization in autobiographical memory. *Memory & cognition* 15, 2 (1987), 119–132.

[21] Martin A Conway and David C Rubin. 2019. The structure of autobiographical memory. In *Theories of memory*. Psychology Press, 103–137.

[22] Masashi Crete-Nishihata, Ronald M Baecker, Michael Massimi, Deborah Ptak, Rachelle Campigotto, Liam D Kaufman, Adam M Brickman, Gary R Turner, Joshua R Steinerman, and Sandra E Black. 2012. Reconstructing the past: personal memory technologies are not just personal and not just for memory. *Human–Computer Interaction* 27, 1-2 (2012), 92–123.

[23] Sara J Czaja, Neil Charness, Arthur D Fisk, Christopher Hertzog, Sankaran N Nair, Wendy A Rogers, and Joseph Sharit. 2006. Factors predicting the use of technology: findings from the Center for Research and Education on Aging and Technology Enhancement (CREATE). *Psychology and aging* 21, 2 (2006), 333.

[24] Zexing Du, Xue Wang, Guoqing Zhou, and Qing Wang. 2022. Fast and unsupervised action boundary detection for action segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 3323–3332.

[25] Be My Eyes. 2023. Be My Eyes Integrates Be My AI™ into its First Contact Center with Stunning Results. https://www.bemyeyes.com/blog/introducing-microsofts-ai-powered-disability-answer-desk-on-be-my-eyes. [Online; accessed 3-March-2024].

[26] Robyn Fivush, Tilmann Habermas, Theodore EA Waters, and Widaad Zaman. 2011. The making of autobiographical memory: Intersections of culture, narratives and identity. *International journal of psychology* 46, 5 (2011), 321–345.

[27] Ana Garcia del Molino, Joo-Hwee Lim, and Ah-Hwee Tan. 2018. Predicting visual context for unsupervised event segmentation in continuous photo-streams. In *Proceedings of the 26th ACM international conference on Multimedia*. 10–17.

[28] Danna Gurari, Qing Li, Chi Lin, Yinan Zhao, Anhong Guo, Abigale Stangl, and Jeffrey P Bigham. 2019. Vizwiz-priv: A dataset for recognizing the presence and purpose of private visual information in images taken by blind people. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 939–948.

[29] Susumu Harada, Daisuke Sato, Dustin W Adams, Sri Kurniawan, Hironobu Takagi, and Chieko Asakawa. 2013. Accessible photo album: enhancing the photo sharing experience for people with visual impairment. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 2127–2136.

[30] Steven CH Hoi, Doyen Sahoo, Jing Lu, and Peilin Zhao. 2021. Online learning: A comprehensive survey. *Neurocomputing* 459 (2021), 249–289.

[31] Ting-Hao Kenneth Huang, Francis Ferraro, Nasrin Mostafazadeh, Ishan Misra, Aishwarya Agrawal, Jacob Devlin, Ross Girshick, Xiaodong He, Pushmeet Kohli, Dhruv Batra, C. Lawrence Zitnick, Devi Parikh, Lucy Vanderwende, Michel Galley, and Margaret Mitchell. 2016. Visual Storytelling. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (San Diego, California), Kevin Knight, Ani Nenkova, and Owen Rambow (Eds.). Association for Computational Linguistics, 1233–1239. https://doi.org/10.18653/v1/N16-1147

[32] Mina Huh, Yunjung Lee, Dasom Choi, Haesoo Kim, Uran Oh, and Juho Kim. 2022. Cocomix: Utilizing Comments to Improve Non-Visual Webtoon Accessibility. In *CHI '22: CHI Conference on Human Factors in Computing Systems, New Orleans, LA, USA, 29 April 2022 - 5 May 2022*, Simone D. J. Barbosa, Cliff Lampe, Caroline Appert, David A. Shamma, Steven Mark Drucker, Julie R. Williamson, and Koji Yatani (Eds.). ACM, 607:1–607:18. https://doi.org/10.1145/3491102.3502081

[33] Mina Huh, Yi-Hao Peng, and Amy Pavel. 2023. GenAssist: Making Image Generation Accessible. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–17.

[34] Apple Inc. 2024. Apple Intelligence. https://www.apple.com/apple-intelligence/. [Online; accessed 30-June-2024].

[35] Ellen Isaacs, Artie Konrad, Alan Walendowski, Thomas Lennig, Victoria Hollis, and Steve Whittaker. 2013. Echoes from the past: how technology mediated reflection improves well-being. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1071–1080.

[36] Jasmine Jones and Mark S. Ackerman. 2018. Co-constructing Family Memory: Understanding the Intergenerational Practices of Passing on Family Stories. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC Canada). ACM, 1–13. https://doi.org/10.1145/3173574.3173998

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

[38] Yunjae Jung, Dahun Kim, Sanghyun Woo, Kyungsu Kim, Sungjin Kim, and In So Kweon. 2020. Hide-and-Tell: Learning to Bridge Photo Streams for Visual Storytelling. 34, 7 (2020), 11213–11220. https://doi.org/10.1609/aaai.v34i07.6780

[39] Jaewook Lee, Jaylin Herskovitz, Yi-Hao Peng, and Anhong Guo. 2022. ImageExplorer: Multi-layered touch exploration to encourage skepticism towards imperfect AI-generated image captions. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–15.

[40] Jaewook Lee, Yi-Hao Peng, Jaylin Herskovitz, and Anhong Guo. 2021. Image Explorer: Multi-layered touch exploration to make images accessible. In *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility*. 1–4.

[41] Matan Levy, Rami Ben-Ari, Nir Darshan, and Dani Lischinski. 2024. Chatting makes perfect: Chat-based image retrieval. *Advances in Neural Information Processing Systems* 36 (2024).

[42] Junnan Li, Dongxu Li, Silvio Savarese, and Steven Hoi. 2023. Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models. In *International conference on machine learning*. PMLR, 19730–19742.

[43] Junnan Li, Dongxu Li, Caiming Xiong, and Steven Hoi. 2022. Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation. In *International conference on machine learning*. PMLR, 12888–12900.

[44] Zhichun Li, Yu Jiang, Xiaochen Liu, Yuhang Zhao, Chun Yu, and Yuanchun Shi. 2022. Enhancing Revisitation in Touchscreen Reading for Visually Impaired People with Semantic Navigation Design. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 3 (2022), 1–22.

[45] Kate Lister, Tim Coughlan, Francisco Iniesto, Nick Freear, and Peter Devine. 2020. Accessible conversational user interfaces: considerations for design. In *Proceedings of the 17th international web for all conference*. 1–11.

[46] Guanhong Liu, Tianyu Yu, Chun Yu, Haiqing Xu, Shuchang Xu, Ciyuan Yang, Feng Wang, Haipeng Mi, and Yuanchun Shi. 2021. Tactile compass: Enabling visually impaired people to follow a path with continuous directional feedback. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–13.

[47] Sonja Lyubomirsky, Kennon M Sheldon, and David Schkade. 2005. Pursuing happiness: The architecture of sustainable change. *Review of general psychology* 9, 2 (2005), 111–131.

[48] David McGookin. 2019. Reveal: Investigating Proactive Location-Based Reminiscing with Personal Digital Photo Repositories. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow Scotland Uk). ACM, 1–14. https://doi.org/10.1145/3290605.3300665

[49] Meredith Ringel Morris, Jazette Johnson, Cynthia L Bennett, and Edward Cutrell. 2018. Rich representations of visual content for screen reader users. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–11.

[50] Quim Motger, Xavier Franch, and Jordi Marco. 2022. Software-based dialogue systems: survey, taxonomy, and challenges. *Comput. Surveys* 55, 5 (2022), 1–42.

[51] Vishnu Nair, Hanxiu 'Hazel' Zhu, and Brian A. Smith. 2023. ImageAssist: Tools for Enhancing Touchscreen-Based Image Exploration Systems for Blind and Low Vision Users. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg Germany, 2023-04-19). ACM, 1–17. https://doi.org/10.1145/3544548.3581302

[52] Theresa Neil. 2014. *Mobile design pattern gallery: UI patterns for smartphone apps*. " O'Reilly Media, Inc.".

[53] Jakob Nielsen. 2005. Ten usability heuristics. (2005).

[54] Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. 2023. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–22.

[55] S. Tejaswi Peesapati, Victoria Schwanda, Johnathon Schultz, Matt Lepage, So-yae Jeong, and Dan Cosley. 2010. Pensieve: supporting everyday reminiscence. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta Georgia USA). ACM, 2027–2036. https://doi.org/10.1145/1753326.1753635

[56] Abhirama Subramanyam Penamakuri, Manish Gupta, Mithun Das Gupta, and Anand Mishra. 2023. Answer Mining from a Pool of Images: Towards Retrieval-Based Visual Question Answering. In *IJCAI*. ijcai.org. https://doi.org/10.24963/ijcai.2023/146

[57] Yi-Hao Peng, Peggy Chi, Anjuli Kannan, Meredith Ringel Morris, and Irfan Essa. 2023. Slide Gestalt: Automatic Structure Extraction in Slide Decks for Non-Visual Access. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–14.

[58] Alisha Pradhan, Kanika Mehta, and Leah Findlater. 2018. "Accessibility Came by Accident": Use of Voice-Controlled Intelligent Personal Assistants by People with Disabilities. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC Canada). ACM, 1–13. https://doi.org/10.1145/3173574.3174033

[59] Emanuele Pucci, Isabella Possaghi, Claudia Maria Cutrupi, Marcos Baez, Cinzia Cappiello, and Maristella Matera. 2023. Defining Patterns for a Conversational Web. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–17.

[60] Woosuk Seo, Chanmo Yang, and Young-Ho Kim. 2024. ChaCha: Leveraging Large Language Models to Prompt Children to Share Their Emotions about Personal Events. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI'24)*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3613904.3642152

[61] 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.

[62] Abigale Stangl, Nitin Verma, Kenneth R Fleischmann, Meredith Ringel Morris, and Danna Gurari. 2021. Going beyond one-size-fits-all image descriptions to satisfy the information wants of people who are blind or have low vision. In *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility*. 1–15.

[63] Jeff Stanley, Ronna ten Brink, Alexandra Valiton, Trevor Bostic, and Becca Scollan. 2022. Chatbot accessibility guidance: a review and way forward. In *Proceedings of Sixth International Congress on Information and Communication Technology: ICICT 2021, London, Volume 3*. Springer, 919–942.

[64] Ella Tallyn, Hector Fried, Rory Gianni, Amy Isard, and Chris Speed. 2018. The ethnobot: Gathering ethnographies in the age of IoT. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–13.

[65] Gemini Team, Rohan Anil, Sebastian Borgeaud, Yonghui Wu, Jean-Baptiste Alayrac, Jiahui Yu, Radu Soricut, Johan Schalkwyk, Andrew M Dai, Anja Hauth, et al. 2023. Gemini: a family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805* (2023).

[66] 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 CHI Conference on Human Factors in Computing Systems*. 1–17.

[67] Oriol Vinyals, Alexander Toshev, Samy Bengio, and Dumitru Erhan. 2015. Show and tell: A neural image caption generator. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3156–3164.

[68] Lei Wang, Chen Ma, Xueyang Feng, Zeyu Zhang, Hao Yang, Jingsen Zhang, Zhiyuan Chen, Jiakai Tang, Xu Chen, Yankai Lin, et al. 2024. A survey on large language model based autonomous agents. *Frontiers of Computer Science* 18, 6 (2024), 1–26.

[69] Weizhi Wang, Li Dong, Hao Cheng, Xiaodong Liu, Xifeng Yan, Jianfeng Gao, and Furu Wei. 2024. Augmenting language models with long-term memory. *Advances in Neural Information Processing Systems* 36 (2024).

[70] Zhan Wang, Lin-Ping Yuan, Liangwei Wang, Bingchuan Jiang, and Wei Zeng. 2024. Virtuwander: Enhancing multi-modal interaction for virtual tour guidance through large language models. In *Proceedings of the CHI conference on human factors in computing systems*. 1–20.

[71] Jeffrey Dean Webster, Ernst T Bohlmeijer, and Gerben J Westerhof. 2010. Mapping the future of reminiscence: A conceptual guide for research and practice. *Research on Aging* 32, 4 (2010), 527–564.

[72] Jing Wei, Sungdong Kim, Hyunhoon Jung, and Young-Ho Kim. 2024. Leveraging Large Language Models to Power Chatbots for Collecting User Self-Reported Data. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1, Article 87 (apr 2024), 35 pages. https://doi.org/10.1145/3637364

[73] Jordan White, William Odom, Nico Brand, and Ce Zhong. 2023. Memory Tracer & Memory Compass: Investigating Personal Location Histories as a Design Material for Everyday Reminiscence. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–19.

[74] Terry Winograd. 1986. A language/action perspective on the design of cooperative work. In *Proceedings of the 1986 ACM conference on Computer-supported cooperative work*. 203–220.

[75] 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.

[76] Tongshuang Wu, Michael Terry, and Carrie Jun Cai. 2022. Ai chains: Transparent and controllable human-ai interaction by chaining large language model prompts. In *Proceedings of the 2022 CHI conference on human factors in computing systems*. 1–22.

[77] Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhudinov, Rich Zemel, and Yoshua Bengio. 2015. Show, attend and tell: Neural image caption generation with visual attention. In *International conference on machine learning*. PMLR, 2048–2057.

[78] Shuchang Xu, Ciyuan Yang, Wenhao Ge, Chun Yu, and Yuanchun Shi. 2020. Virtual Paving: Rendering a smooth path for people with visual impairment through vibrotactile and audio feedback. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 3 (2020), 1–25.

[79] Ciyuan Yang, Shuchang Xu, Tianyu Yu, Guanhong Liu, Chun Yu, and Yuanchun Shi. 2021. LightGuide: Directing Visually Impaired People along a Path Using Light Cues. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 2 (2021), 1–27.

[80] Pengcheng Yang, Fuli Luo, Peng Chen, Lei Li, Zhiyi Yin, Xiaodong He, and Xu Sun. 2019. Knowledgeable Storyteller: A Commonsense-Driven Generative Model for Visual Storytelling. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence* (Macao, China, 2019-08). International Joint Conferences on Artificial Intelligence Organization, 5356–5362. https://doi.org/10.24963/ijcai.2019/744

[81] Zhengyuan Yang, Linjie Li, Kevin Lin, Jianfeng Wang, Chung-Ching Lin, Zicheng Liu, and Lijuan Wang. 2023. The dawn of lmms: Preliminary explorations with gpt-4v (ision). *arXiv preprint arXiv:2309.17421* 9, 1 (2023), 1.

[82] MinYoung Yoo, William Odom, and Arne Berger. 2021. Understanding Everyday Experiences of Reminiscence for People with Blindness: Practices, Tensions and Probing New Design Possibilities. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama Japan). ACM, 1–15. https://doi.org/10.1145/3411764.3445212

[83] Youngjae Yu, Jiwan Chung, Heeseung Yun, Jongseok Kim, and Gunhee Kim. 2021. Transitional Adaptation of Pretrained Models for Visual Storytelling. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (Nashville, TN, USA, 2021-06). IEEE, 12653–12663. https://doi.org/10.1109/CVPR46437.2021.01247

[84] Yuhang Zhao, Shaomei Wu, Lindsay Reynolds, and Shiri Azenkot. 2017. The effect of computer-generated descriptions on photo-sharing experiences of people with visual impairments. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 1–22.

[85] Wanjun Zhong, Lianghong Guo, Qiqi Gao, He Ye, and Yanlin Wang. 2024. Memorybank: Enhancing large language models with long-term memory. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38. 19724–19731.

[86] Yu Zhong, Walter S Lasecki, Erin Brady, and Jeffrey P Bigham. 2015. Regionspeak: Quick comprehensive spatial descriptions of complex images for blind users. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 2353–2362.

A TABLES

Table 3 shows the demographics of the participants. Table 4 shows the subjective ratings in the evaluation study.

Table 3: Demographic information of the participants.

PID	Age	Gender	Visual Condition	Onset	Image Accessibility Tools	Chatbot Usage	Themes of Photos in Study
P1	29	M	Legally blind	Acquired	Seeing AI, Talkback	Huawei Xiaoyi	Family gatherings
P2	45	F	Totally blind	Congenital	VoiceOver	Siri	Family trips
P3	26	M	Totally blind	Acquired	Be My AI, VoiceOver	ChatGPT, Siri	Family trips
P4	27	F	Legally blind	Congenital	Be My AI, VoiceOver	Siri	Stage performances
P5	38	M	Totally blind	Congenital	Be My AI, VoiceOver	Siri	Family gatherings
P6	35	F	Totally blind	Congenital	Talkback	Xiaomi xiao'ai	Ceremonies
P7	30	F	Totally blind	Congenital	Be My AI, VoiceOver	ChatGPT, Siri	Business trips
P8	33	M	Legally blind	Congenital	Be My AI, Talkback	Xiaomi xiao'ai	Trips with friends
P9	52	F	Totally blind	Acquired	VoiceOver	Siri	Family gatherings
P10	30	F	Totally blind	Congenital	Be My AI, VoiceOver	Siri	Trips abroad
P11	32	M	Legally blind	Acquired	Be My AI, Talkback	ChatGPT	Trips abroad
P12	31	M	Totally blind	Congenital	Be My AI, VoiceOver	Gemini	Conference speech
P13	44	F	Legally blind	Acquired	Be My AI, VoiceOver	Siri	Family trips
P14	36	M	Legally blind	Congenital	VoiceOver	ChatGPT, Siri	Family trips
P15	27	F	Legally blind	Congenital	VoiceOver	Siri	Trips with friends
P16	26	M	Legally blind	Congenital	Be My AI, VoiceOver	Siri	Stage performances
P17	30	F	Totally blind	Acquired	Be My AI, VoiceOver	ChatGPT, Siri	Family trips
P18	36	M	Totally blind	Congenital	Be My AI, VoiceOver	Siri	Ceremonies

Table 4: Subjective ratings. (1=Strongly Disagree, 7=Strongly Agree. Wilcoxon signed-rank test was used for significance analysis.)

Aspects	Participant statements	Memory Reviver	Baseline	Significance
Reminiscence Experience	**Effectiveness**: I fully recalled memories about past events.	6.75 (SD=0.45)	4.42 (SD=1.08)	$Z = -3.08, p < .01$
	Enjoyment: I felt happy when reminiscing with this chatbot.	6.17 (SD=0.83)	4.58 (SD=1.56)	$Z = -2.70, p < .01$
	Low Mental Demand: I didn't feel mentally demanded using this chatbot.	6.75 (SD=0.45)	5.00 (SD=1.60)	$Z = -2.68, p < .01$
	Overall Satisfaction: I am satisfied with the reminiscence experience.	6.33 (SD=0.49)	4.50 (SD=0.80)	$Z = -3.17, p < .01$
Understanding a Collection	**Clear Storyline**: I clearly grasped the storyline of all the scenes.	6.75 (SD=0.45)	4.42 (SD=0.90)	$Z = -3.11, p < .01$
	Easy Recall of Activities: I easily recalled past activities in each scene.	6.83 (SD=0.39)	4.75 (SD=1.29)	$Z = -2.88, p < .01$
	Easy Exploration of Details: I easily explored the details in each scene.	6.83 (SD=0.39)	3.58 (SD=1.31)	$Z = -3.07, p < .01$
	Easy Discovery of New Scenes: I easily found new scenes to talk about.	6.50 (SD=0.67)	4.58 (SD=1.68)	$Z = -2.97, p < .01$
Conversational Experience	**Natural Conversation Flow**: The conversation flowed naturally.	6.58 (SD=0.67)	5.92 (SD=1.31)	$Z = -1.63, p = 0.10$
	Reply Satisfaction: The chatbot provided satisfying replies.	6.50 (SD=0.52)	4.75 (SD=1.54)	$Z = -2.54, p < .01$
	Human-likeness: I felt like I was talking to a real person.	5.50 (SD=1.51)	3.75 (SD=2.09)	$Z = -2.70, p < .01$