Eye2Recall: Exploring Mixed-Initiative Reminiscence Activities via Gaze-Driven LLM Prompts for Older Adults

Eye2Recall: Fusing Gaze and LLMs for Mixed-Initiative Reminiscence with Older Adults

Lei Han
Computational Media and Arts Thrust
Hong Kong University of Science and Technology (Guangzhou)
Guangzhou, China
lhan229@connect.hkust-gz.edu.cn

Mingnan Wei
Bioscience and Biomedical Engineering
The Hong Kong University of Science and Technology (Guangzhou)
Guangzhou, China
mwei881@connect.hkust-gz.edu.cn

Qiongyan Chen
Computational Media and Arts
The Hong Kong University of Science and Technology (Guangzhou)
Guangzhou, China
qchen580@conncet.hkust-gz.edu.cn

Anqi Wang
Emerging Interdisciplinary Areas
Hong Kong university of Science and Technology
Hong Kong, China
Computational Media and Arts
Hong Kong University of Science and Technology(Guangzhou)
Guangzhou, China
awangan@connect.ust.hk

Rong Pang
Edinburgh College of Art
University of Edinburgh
Edinburgh, United Kingdom
rong.pang2024@outlook.com

David Yip*
Computational Media and Arts
The Hong Kong University of Science and Technology (Guangzhou)
Gaungzhou, China
daveyip@hkust-gz.edu.cn

Abstract

Photo-based reminiscence can support well-being in older adults, yet most systems remain text-driven and offer little real-time adaptivity. We first conduct expert interviews to derive design considerations for accessibility, cultural fit, and safe emotional engagement. We then implemented *Eye2Recall*, an intelligent conversational interface that converts users' gazes on old photos into mixed-initiative prompts for a large language model (LLM). We evaluated it in a pilot study with 12 older adults. Participants reported low-effort, smooth interactions, and perceived the agent's questions as aligned with what they were looking at. Immediately after use, self-reported positive mood increased and negative mood decreased. Interviews further indicated that gaze-driven prompts helped retrieve concrete details and supported reflective storytelling. Our contribution is a concrete mechanism for gaze-to-prompt adaptivity that operationalizes mixed-initiative dialogue for older adults' reminiscence experience.

CCS Concepts

• **Human-centered computing → User centered design**.

*Corresponding Author

ACM ISBN 979-8-4007-1984-4/26/03
https://doi.org/10.1145/3742413.3789085

Keywords

Older Adult; Human-AI Interaction; Eye-tracking; Digital Reminiscence; Positive Aging.

ACM Reference Format:
Lei Han, Mingnan Wei, Qiongyan Chen, Anqi Wang, Rong Pang, and David Yip. 2026. Eye2Recall: Exploring Mixed-Initiative Reminiscence Activities via Gaze-Driven LLM Prompts for Older Adults: Eye2Recall: Fusing Gaze and LLMs for Mixed-Initiative Reminiscence with Older Adults. In *31st International Conference on Intelligent User Interfaces (IUI '26), March 23–26, 2026, Paphos, Cyprus.* ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3742413.3789085

1 Introduction

Reminiscence is considered an effective approach for improving the well-being of older adults [10, 52], positively impacting emotional health for both healthy older adults and individuals with dementia [14, 25]. In addition, reminiscence can enhance communication skills and strengthen intergenerational relationships [26, 46, 87].

In recent decades, research has increasingly explored how technology can support reminiscence and enrich older adults' daily lives. Various digital tools have been developed to evoke memories through interactive engagement with photos [37, 63, 87], sounds [19, 34, 35], and other sensory stimuli. With the advancement of large language models (LLMs), studies have investigated LLM-powered solutions for digital reminiscence, including collecting personal memories [62], supporting storytelling [15], and facilitating reminiscence conversations [88].

Despite these advances, existing technologies still face challenges in effectively supporting older adults' reminiscence [90, 93]. Many LLM-powered conversational agents (CAs) rely primarily on text

Figure 1: Participants use the *Eye2Recall* prototype to interact with the LLM-powered agent.

or speech inputs, which may create barriers due to complex user interfaces or literacy requirements [77, 90]. Additionally, these systems often lack personalization and struggle to adapt dynamically to users' needs, while speech-based agents face difficulties understanding older adults' accents and nuances, potentially reducing conversational coherence [36, 57].

Previous studies suggest that eye-tracking technology can help older adults express their needs and preferences more effectively, particularly for those with communication difficulties [48, 61]. Recent HCI and AI research has employed gaze information to guide AI models by aligning attention with human fixations and areas of interest [49, 69, 82, 84, 89]. However, little is known about using eye-tracking regions of interest (ROI) as cues for LLMs to support gaze-driven reminiscence dialogues with older adults. Gaze information reflects pre-intentional attention dynamics and implicit cues [33, 66], which can support nuanced and unstructured interactive activities, such as reminiscence. Leveraging this capability, gaze information allows conversational systems to adapt dynamically to users' attention and interests, enhancing personalization and engagement in the such experiences. In this context, we propose two research questions:

- **RQ1:** *How can the integration of LLM with eye-tracking technology enhance the conversational experience for older adults during reminiscence activities?*
- **RQ2:** *What effects does LLM-powered conversation with eye-tracking augmentation have on older adults' workload, usability, and affect in photo-based reminiscence?*

To address these questions, we conducted a semi-structured interview study with experts to understand older adults' needs and preferences during photo-based reminiscence. Based on the findings, we designed *Eye2Recall*, an LLM-powered prototype that integrates eye-tracking with visual interest detection to adapt dialogue content in real time (see Fig. 1). To evaluate the prototype and derive design implications, we conducted a user study (N=12) aged 60 and above, including pre-evaluation, prototype testing, semi-structured interviews, and post-evaluation, to comprehensively investigate older adults' reminiscence experiences with a gaze-augmented conversational system. In summary, this work makes three key contributions:

- Conducted expert interviews to identify design considerations, challenges, and opportunities for LLM tools that support photo-based reminiscence with older adults.
- Developed *Eye2Recall*, a prototype that uses gaze-based cues to personalize LLM-mediated reminiscence conversations around old photos.
- Ran a user study with older adults to evaluate *Eye2Recall*'s usability, workload, and user experience, yielding design insights and implications for future LLM-supported reminiscence systems.

2 Related Work

2.1 Psychological Support with Reminiscence

Reminiscence, which involves recalling and sharing past experiences, has long been recognized as a meaningful way to promote psychological well-being in older adults [10, 52]. It can enhance mood, foster positive emotions, and build psychological resilience [24, 27]. Such recollection is often triggered by sensory cues, such as photographs or music, which evoke autobiographical memories and shape corresponding emotional responses [3, 44, 67, 80]. However, access to human-facilitated reminiscence remains uneven: caregivers' time is limited and professional services are costly, placing emotional and logistical burdens on families and care systems [50].

These constraints highlight the need for scalable, low-effort digital tools that can support memory elicitation in everyday contexts.

Beyond short-term emotional outcomes, reminiscence also strengthens self-identity and a sense of belonging by prompting reflection on life accomplishments and enabling the sharing of personal stories within social contexts [9, 42]. As a therapeutic approach, reminiscence and narrative techniques have been used to engage people living with depression, dementia, or cognitive decline, often producing measurable improvements in emotional health [11, 25, 72]. Among different modalities, photo-elicited reminiscence has proven particularly effective: old images can cue autobiographical memory, help retrieve self-defining experiences, and reinforce a coherent sense of life continuity [18, 64, 79, 81]. This body of work motivates designing technologies that make photo-based reminiscence more accessible, adaptive, and emotionally supportive for older adults.

2.2 Conversational Systems for Reminiscence

Recent HCI research has increasingly explored how conversational agents (CAs) can facilitate reminiscence for older adults [4, 70, 73, 85]. Advances in natural language processing (NLP) and machine learning (ML) have enabled CAs capable of engaging users in fluid, human-like dialogue [39]. For example, *GoodTimes* is an interactive photo album application that uses AI to converse about personal pictures and life events, promoting engagement around family and social memories [86]. Such systems are commonly categorized as either task-oriented or open-domain dialogue systems [28, 32].

Task-oriented systems are often designed for therapeutic goals such as emotional regulation or cognitive intervention [35, 60]. *Yeonheebot*, for instance, provides automated reminiscence therapy to mitigate depression and anxiety in older adults [68], demonstrating the potential of conversational technologies in dementia care and mental health contexts [40]. Other efforts have used NLP and ML to analyze older adults' natural conversation patterns and promote social reminiscence [23, 92]. However, many of these systems remain rule-based or rely on predefined conversation flows, which limit their ability to handle unexpected topics or follow users' emotional trajectories [41, 51, 78].

Open-domain dialogue systems offer greater flexibility, supporting natural and wide-ranging conversation. Studies have shown that open-ended questions can improve autobiographical recall [29], but generating such prompts in a way that feels empathetic and contextually appropriate remains a key challenge. The rapid advancement of large language models (LLMs) has transformed this space: LLM-powered systems can generate coherent, context-aware dialogue through in-context learning [70]. For example, Purohit et al. demonstrated how LLM-based methods can aid word retrieval for individuals with aphasia, suggesting broader applications in reminiscence therapy [65]. Yet, systematic exploration of LLM-powered dialogue for older adults' reminiscence remains limited. Most current systems neglect the subtle cues—such as attention or engagement—that could help adapt conversation pacing and content in real time [6]. This limitation motivates exploring new modalities for implicit interaction and personalization.

2.3 Eye-Tracking as Alternative Input Modality

Gaze-driven interaction has long been studied as a means of capturing users' attentional processes [69, 89]. More recent work has incorporated eye movements into language-based modeling to reflect user interest in specific image regions and to guide AI attention and prediction[49]. Unlike explicit cues such as pointing or verbal references, which require users to externalize well-formed intentions, gaze reflects pre-intentional attention dynamics that emerge during ongoing perception and sense-making [2, 22, 33, 38, 66]. Lopez-Cardona et al. propose *GazeReward*, which incorporates implicit eye-tracking feedback into an LLM's reward model, showing that gaze signals significantly improve alignment with human preferences [49]. Other architectures encode gaze as structured attention dynamics. For example, STARE tokenizes spatio-temporal eye-movement data over predefined regions of interest (ROIs) and feeds these representations into a transformer model to predict user choices [84]. Likewise, the *Gaze2AOI* automatically identifies semantically meaningful image regions and links these Areas of Interest (AOIs) with gaze metrics (e.g. fixation dwell time), illustrating how ROIs can be inferred directly from eye fixations [82]. These works collectively demonstrate that eye movements reflect user interest in specific image regions and that integrating gaze information can guide AI attention and predictions.

Beyond serving as an alternative input modality, gaze-driven prompts differ fundamentally from explicit cues such as pointing or verbal referencing [31, 66]. While deictic gestures and linguistic references require users to externalize well-formed intentions, gaze operates as an implicit and pre-linguistic signal that reflects emergent attention dynamics during ongoing perception and sense-making. As such, gaze captures cognitive traces before they are stabilized into explicit intent or symbolic expression, rather than deliberate communicative acts [2, 22, 38].

This implicit property makes gaze particularly effective for supporting reminiscence, which relies on revisiting prior attentional and interpretive states rather than recalling fully articulated content. Cognitive studies on episodic memory retrieval show that eye movements often reenact original attention patterns even in the absence of visual stimuli, a phenomenon known as the "looking-at-nothing" effect [67, 83]. By externalizing attention history, gaze provides access to latent, non-verbal memory cues that are difficult to verbalize yet critical for reflective reconstruction. Consequently, gaze-driven prompts enable AI systems to scaffold reminiscence by aligning with how users previously attended, interpreted, and assigned meaning, rather than solely responding to post-hoc verbal descriptions [49, 89].

3 Expert Interview

We conducted expert interviews with four domain experts to explore older adults' nostalgic characteristics and the potential challenges and opportunities they may face when participating in reminiscence activities. Then, we outlined the design considerations and potential challenges of constructing an AI-supported reminiscence system for older adults.

Table 1: Overview of Expert Participants

ID	G	Role	Affiliation	Expertise
E1	F	Professor	University	Gerontology; Cognitive psychology
E2	M	Professor	University	Computer science; HCI
E3	F	Social worker	Public social services	Older-adult counseling; Memory rehabilitation
E4	M	Professor	University	Visual neuroscience

3.1 Participants and Procedure

We recruited four experts from local universities and public service agencies for one-to-one, online semi-structured interviews. The panel comprised three professors with relevant research expertise and one social worker with direct practice experience with older adults (see Table 1).

Interviews began with background questions about each expert's role and career, followed by prompts on how AI might support old-photo–based digital reminiscence for older adults and the challenges older adults may encounter in this process. Each session lasted approximately 30 minutes (see Appendix F). With permission, all interviews were recorded.

3.2 Data Analysis

We conducted an inductive thematic analysis [54] of the expert interviews. Two researchers independently coded the Chinese transcripts and iteratively developed a shared codebook. After calibrating on 20–30% of the data, we refined coding rules and applied the finalized codebook to the full corpus. Inter-rater agreement on the calibrated subset was substantial (Cohen's κ = .72); disagreements were resolved by consensus and logged. For bilingual reporting, themes and exemplar quotes were translated into English and verified by a second author, with contentious cases resolved via discussion and selective back-translation.

3.3 Design Considerations

Based on our findings, we propose two design considerations (DCs) for AI-assisted reminiscence using old photos with older adults. Each DC articulates a rationale and translates it into actionable design rules, supported by participant evidence.

3.3.1 DC1: Low-Effort, Accessible, and Safe Mixed-Initiative Reminiscence.

- **Keep interaction natural with minimal cognitive load.** Experts ($E2, E4$) emphasized simplicity and clarity, noting that multi-step, browser-style controls can disrupt the train of thought. The interface should prioritize gaze and speech as the primary modalities and avoid nested menus. As $E2$ put it, *"Some browser-based interfaces pose challenges for older adults and may break their train of thought."*
- **Provide guided, empathetic AI facilitation.** Experts ($E1$, $E2$) reported that encouraging prompts may help older adults articulate stories and emotions. The dialogue agent should acknowledge and probe while pacing the conversation (e.g., reflective paraphrasing), and should avoid jargon or rapid-fire questioning. As $E1$ noted, *"Reminiscence aims to evoke personal memories through external memory cues, so the dialogue structure should be inspiring."*
- **Ensure accessibility by default.** Experts ($E2, E4$) highlighted practical accommodations, including high-contrast visuals, large text and targets, clear audio with adjustable volume, and minimal on-screen controls. Where available, a large display can improve legibility and shared viewing ($E2$). Echoing this, $E4$ emphasized that the UI should remain simple and senior-friendly, with reduced visual clutter and clearly distinguishable controls (e.g., larger buttons and straightforward layouts).
- **Convey safety and privacy to build trust.** Experts stressed the need to communicate what is recorded and why, and to keep data secure. The system should provide plain-language notices, allow pausing/opt-out at any time, and store logs locally and/or in anonymized form. As $E3$ remarked, *"It is vital to ensure older adults understand the system is secure and will not compromise their privacy."*

3.3.2 DC2: Content Strategy: Cultural Fit, Effective External Cues, and Chronological Structuring.

- **Cultural background.** Experts ($E2$, $E3$) recommended selecting photos aligned with local traditions and history so that cues felt familiar and meaningful. As $E3$ noted, *"Selecting old photos that reflect local traditions and history can help them relive and appreciate past experiences and culture."* $E2$ added, *"Some existing reminiscence tools provide content that local older adults find difficult to relate to."*
- **External memory cues.** Beyond personal albums, era-typical photos can serve as effective cues that stimulate curiosity and enrich storytelling, complementing personal recollection. As $E1$ suggested, era-specific collective photos *"can effectively trigger memories even if older adults have never seen the exact photo before."* Prior work has also shown that generic old photos used as external cues can support reminiscence outcomes for older adults, including those with dementia [8, 11, 91].
- **Chronological order.** Experts ($E1$, $E3$) recommended organizing selected photos along life-course milestones (e.g., childhood, early adulthood, marriage, parenting, career, retirement) so narratives flowed coherently and context was easier to retrieve ($E3$). Such structuring can reduce cognitive load by providing a predictable narrative scaffold, helping older adults orient themselves in time and smoothly transition between episodes.

4 System Implementation

Guided by our design considerations (Section 3.3) from expert interviews, we developed *Eye2Recall*, a prototype AI conversational reminiscence system that supports older adults sharing memories (see Fig. 2). The prototype comprises three components: ***Content Design***, ***Gaze-based ROI Detection***, and ***Gaze-to-Prompt Adaptation***.

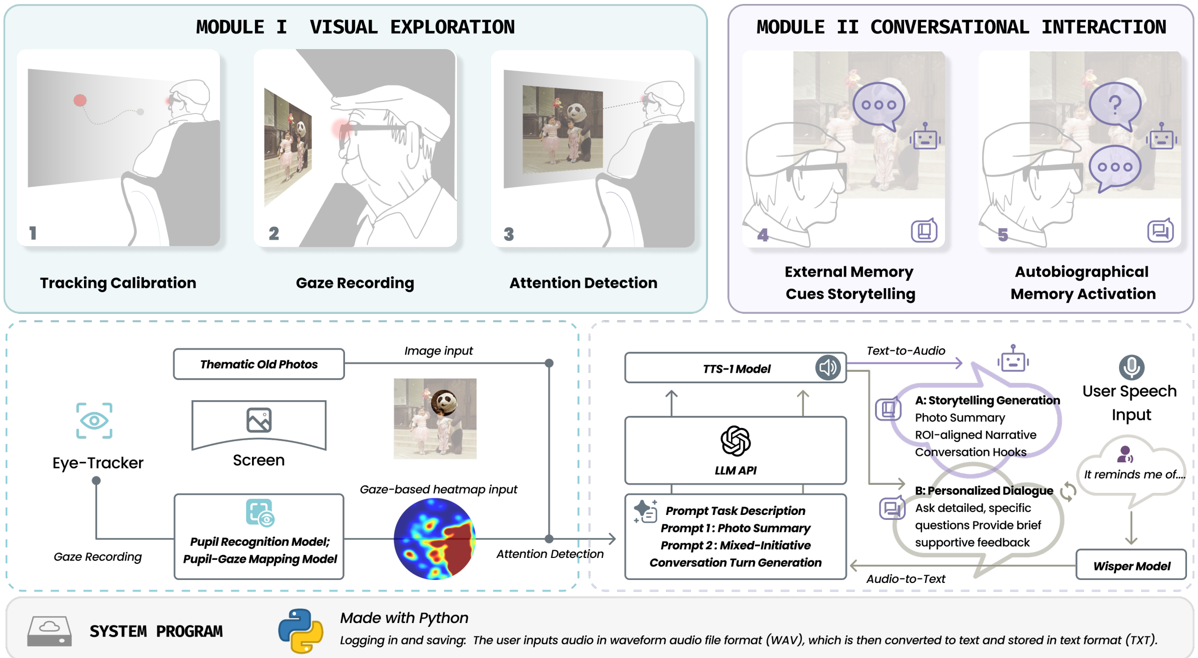


Figure 2: System pipeline of the *Eye2Recall* prototype. The system includes (I) a visual exploration module with eye-tracking calibration, gaze recording, and attention/ROI detection, and (II) a conversational interaction module comprising external-cue–guided storytelling and autobiographical memory activation. Steps 2–5 repeat for each user query in our LLM-powered pipeline.

Figure 3: (a) Example of visual content categorization, consisting of generic old photos categorized into five themes. (b) Screenshots of the *Eye2Recall* user interface (UI).

4.1 Content Design

Guided by **DC2**, we curated the visual content to align with users' cultural backgrounds, incorporate external memory trigger cues, and present materials in a chronological sequence. In parallel, we designed the user interface to surface these cues clearly and support low-effort, timeline-based navigation.

4.1.1 Reminiscence Material. Considering local older adults' cultural context, we curated approximately 2,000 era-typical archival photographs from publicly accessible collections [47, 59, 76]. The corpus spans the 1970s–1990s and emphasizes everyday life, public events, and locally salient scenes. We applied inclusion criteria (Chinese context; usable resolution; basic metadata such as year or location) and excluded images with sensitive content or clearly identifiable individuals without permission. These photographs serve as external memory cues for photo-elicited reminiscence, providing culturally familiar triggers to support recall and engagement.

To support chronology while retaining meaningful cues, we adopted a two-step organization: (1) theme tagging informed by Boyer's notion of collective memory in the city and place-based perspectives on memory [12, 75]; and (2) within each theme, chronological ordering by available metadata (year/period). In line with reminiscence research on external cues [3], the era-typical photos were grouped into five themes with operational inclusion criteria: (i) *Childhood*: early-life settings and relations (home, school, play, caregivers). (ii) *Cultural Heritage*: locally salient symbols, festivals, and folk practices (parades, rites, crafts). (iii) *Urban Development*: urban change and everyday work contexts (streetscapes, housing, transport, industry). (iv) *Migration & Mobility*: leaving/arriving, travel and relocation for work, study, or life transitions. (v) *Life Events*: personal milestones (birthdays, weddings, graduations) depicted in public or family spaces. Representative examples for each theme are shown in Fig. 3a.

4.1.2 User Interface. To ensure accessibility and immersion [17], we selected photos with clear, salient visual cues (e.g., adequate sharpness and colour, discernible depth and perspective, and coherent spatial layout and motion direction). The user interface displays the historical photo centrally on a large LED screen, while removing non-essential on-screen elements and using a clean, uniform background (see Fig. 3b). This layout keeps the visual focus on the photo content and reduces potential distractions during conversation.

4.2 Gaze-based ROI Detection

Guided by **DC1**, in the *Visual Exploration Module* (see Fig. 2), we use a glasses-based eye tracker to detect users' visual interests while users view each photo. During viewing, the prototype samples and derives gazes (location and duration) and saccades. Gazes are identified using a duration threshold (e.g., ≥300 ms) and clustered on the photo plane to obtain ROIs. From these data, the prototype renders gaze-based heatmaps that summarize the distribution of visual attention (see Appendix C). The gaze-based heatmaps are then passed to the *Conversational Interaction Module* to condition the next mixed-initiative turn.

4.3 Gaze-to-Prompt Adaptation

We configured the agent with two tailored prompt instructions (see Appendix A for more details). Leveraging the content of the current photo and the user's gaze-based heatmap, these prompts drive personalized, context-aware dialogue turns.

Prompt 1 provides contextual scaffolding by instructing the agent to **summarize** the current photo with brief background details to open the conversation (see Fig. 4). In pilot tuning, we set the generation parameters to temperature = 1.0 and a response length of approximately 600 tokens, yielding 1–2 min spoken summaries and a good balance between breadth and relevance.

Prompt 2 conditions the agent on the **gaze-based heatmap** to ask **targeted, personalized** questions about the top ROIs, combining external cues with autobiographical recall, hard-capped at two questions per turn (see Fig. 4). Parameters were set to temperature = 0.5 and approximately 200 tokens, yielding concise 20–40 s follow-ups that maintained attention without overloading users.

Text outputs were synthesized to speech with TTS-1 and presented to the user. Each photo was handled in two dialogue turns (Prompt 1 followed by Prompt 2), after which the system automatically advanced to the next photo. Conversation transcripts between the agent and the participant are shown in Fig. 4.

4.4 Implementation Details

The hardware setup includes a glasses-based eye tracker, a large LED display, external stereo speakers, a wireless lapel microphone, and a host workstation for synchronized data capture. The software stack uses a vision-capable LLM API[1] for image understanding and prompt generation, a text-to-speech model [2] for spoken output, and an automatic speech recognition model [3] for transcribing participants' speech. Detailed eye-tracker specifications and calibration procedures are provided in Appendix B. All data were de-identified and stored on an encrypted local drive; personally identifying information was removed after use. API calls were configured for transient processing only, and local session logs can be purged at the end of each study session.

5 User Study

We conducted a user study (N=12) to evaluate the *Eye2Recall*, focusing on usability, workload, engagement, and affective outcomes. The study design and procedure are summarized in Fig. 5. The study protocol was reviewed and approved by the institutional ethics committee of HKUST (Guangzhou) (Approval No. HKUST(GZ)-HSP-2025-0394), and all participants provided informed consent and could withdraw at any time without penalty. We collected only study-relevant data (questionnaires, interaction logs, and interviews); personally identifiable information was not stored with the research data. All datasets were pseudonymized using participant codes, stored on encrypted drives with access restricted to the research team, and reported only in aggregate form or through anonymized quotations. A qualified social worker was available on call to support participants and to implement a predefined distress-and-pausing procedure when needed.

5.1 Participants

We recruited a sample of 12 participants (5 male and 7 female) through online social media recruitment (see Table 2). Participants were aged between 60 and 84 years ($M = 70.42, SD = 6.08$). Participants reported normal or corrected-to-normal vision and hearing, and had no motor impairments that would prevent operation of the study apparatus. Each participant received 200 RMB (about 27 USD) as compensation.

[1]GPT-4-vision-preview API Documentation. Available at: https://platform.openai.com/docs/guides/vision/vision

[2]TTS-1 API Documentation. Available at: https://platform.openai.com/docs/models/tts-1

[3]Whisper API Documentation. Available at: https://platform.openai.com/docs/guides/speech-to-text

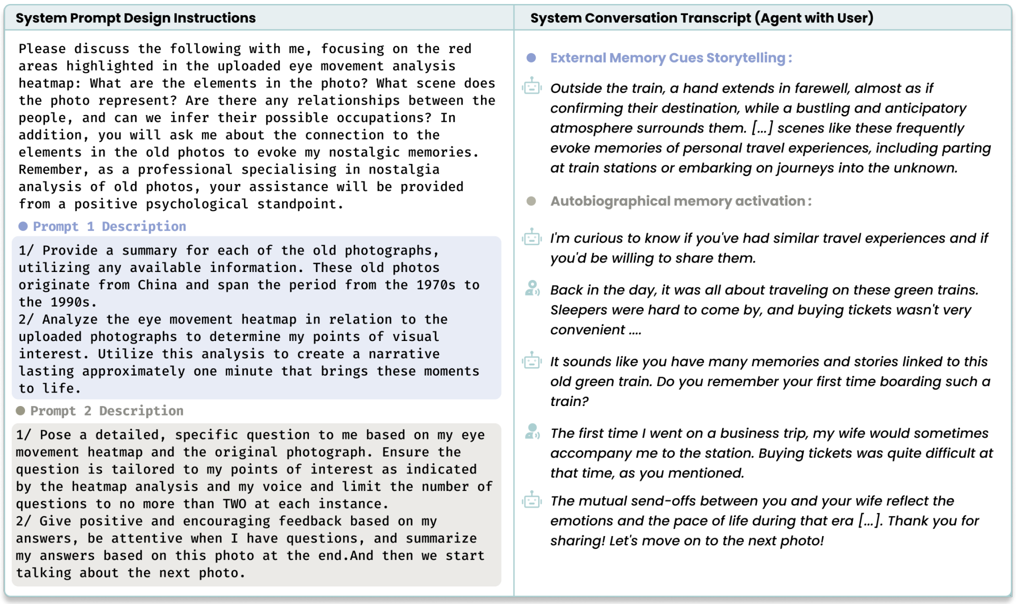


Figure 4: Prompt task instructions and Chat History in *Eye2Recall*. On the left side, there are descriptions outlining the overall task along with detailed instructions for two distinct modules. On the right side, an example conversation between an LLM-powered agent and a user is provided.

Table 2: Demographics of Older Adult Participants.

ID	Gender	Age	Education Level	MMSE
P1	M	73	Teacher's training college	30
P2	F	73	Teacher's training college	29
P3	F	64	High school	29
P4	M	84	High school	27*
P5	F	67	High school	30
P6	M	70	Teacher's training college	29
P7	F	70	University	29
P8	F	60	High school	27*
P9	M	76	High school	30
P10	M	67	University	30
P11	F	69	University	28
P12	F	72	High school	29

Note. *MMSE scores for P4 and P8 are below the normal range, indicating potential risk of mild cognitive impairment (MCI).

According to the Chinese version of the *Mini–Mental State Examination (MMSE)* [43] and our screening protocol, scores of 28–30 were treated as within the normal range, whereas scores $\leq$ 27 suggested a risk of mild cognitive impairment (MCI); thresholds may vary by age and education, and the MMSE is a screening tool rather than a diagnostic test. Across all participants (M = 28.92, SD = 1.08), 10 screened within the normal range and 2 screened at risk of MCI. No participant scored < 21 under our protocol; thus, no case met our criterion for severe impairment risk.

5.2 Evaluation Dimensions

5.2.1 Quantitative Evaluation. We adopted three key dimensions for quantitative evaluation: Perceived usability, Affect and Task Performance.

The **Perceived usability** dimension captured participants' subjective experiences of usability and task load. To measure this, we used a custom user experience questionnaire and the *NASA Task Load Index* (NASA-TLX) [30], both rated on a 7-point Likert scale. The custom questionnaire assessed *accessibility, effectiveness, and overall user experience* (see Appendix G), adapted from the *System Usability Scale* (SUS) [13].

The **Affect** dimension captured participants' emotional state prior and post to system use, measured with the *Positive and Negative Affect Schedule* (PANAS); PA and NA subscales [16]. This scale, comprising 20 items, equally measures positive affect (e.g., *excited, inspired*) and negative affect (e.g., *upset, anxious*) and is well-regarded for capturing emotional nuances [71]

Interaction dynamics. We analyzed eye-tracking data for all participants, including gaze-based heatmaps, gaze ratios, saccade frequencies, response duration and response latency. Two researchers examined gaze distributions to identify common and divergent gaze patterns across participants. We applied a descriptive, mixed-methods approach to the conversational content. We used the text-mining platform *Weiciyun*[4] to conduct word-frequency [7, 56].

5.2.2 Qualitative Evaluation. To complement these quantitative methods, we conducted a **semi-structured interview** to gather qualitative feedback on participants' experiences, providing deeper insights into strengths, weaknesses, and improvement areas. All recordings were transcribed using a commercially available automatic speech recognition (ASR) system [5]. Two researchers performed thematic analysis [54] with an iterative codebook. Disagreements were resolved through discussion until consensus. The

[4] *Weiciyun*, Available at https://www.weiciyun.com.
[5] *iFLYTEK*. Available at https://www.iflyrec.com/zhuanwenzi.html

qualitative results were used to explain and extend the quantitative findings, with particular attention to concrete moments where gaze-driven prompts supported reminiscence or caused friction.

5.3 Procedure

The study was conducted by two researchers, with a professional social worker available on call, in a university laboratory. Each session lasted approximately 1.5 hours and followed three phases (see Fig. 5).

5.3.1 Introduction and pre-evaluation. We introduced the system and procedures, obtained written informed consent, and administered a demographics form. The Chinese version of MMSE was used to screen cognitive status. At the beginning of the session, we assessed participants' alertness using the *Stanford Sleepiness Scale* (SSS) [74]. Pre-session affect was measured using the PANAS. To reduce burden, all items were presented in large fonts with high contrast. The researcher could read items aloud on request.

5.3.2 Prototype testing. The test phase lasted approximately 50 minutes and followed a two-stage workflow—*Visual Exploration* and *Conversational Interaction*—as summarized below.

- ***Visual Exploration*** : (i) Participants were seated comfortably, fitted with the glasses-based eye tracker, and completed a standard calibration. They then rested for about one minute with eyes closed (see Fig. 6(a)). (ii) Participants viewed five era-typical photos on the LED display, 1 minute per photo, in a predetermined order while an experimenter supervised the procedure (see Fig. 6(b)).
- ***Conversational Interaction***: (i) The agent delivered a brief spoken summary (*Prompt 1*) of the current photo (approximately 1–2 minutes) via external speakers (see Fig. 6(c)). (ii) Using gaze-based heatmap, the agent then asked up to two targeted questions about each photo (*Prompt 2*), and participants responded via a wireless microphone. After two dialogue turns per photo, the system automatically advanced to the next photo.

5.3.3 Post-evaluation and semi-structured interview. After the test phase, participants completed the custom user experience questionnaire, PANAS and NASA–TLX. A semi-structured interview (approximately 30 minutes) was then conducted to elicit perceptions of usability, workload, pacing, prompt relevance, and suggestions for improvement.

6 Findings

This user study aimed to evaluate the older adult component of *Eye2Recall*. Overall, the results indicate that the pilot system demonstrated good usability (Section 6.1), enabled older adults to actively engage in LLM-mediated conversations through gaze-based interaction (Section 6.2), and provided perceived emotional and well-being support (Section 6.3).

6.1 Impact on Perceived Usability

6.1.1 NASA-TLX Questionnaire. We measured participants' perceived **workload** using the NASA–TLX. Overall workload was low ($M = 1.74$, $SD = 0.28$), all six sub-scale means were well below the midpoint (see Fig. 7), indicating consistently low perceived workload. By subscale, *Mental Demand* had the highest mean ($M = 2.42$, $SD = 0.86$), which suggest that participants invested cognitive effort in interpreting photos and formulating narratives. It was followed by *Effort* ($M = 1.92$, $SD = 0.76$). *Temporal Demand* ($M = 1.58$, $SD = 0.64$) and *Performance* ($M = 1.58$, $SD = 0.76$) were lower, while *Physical Demand* ($M = 1.50$, $SD = 0.65$) and *Frustration* ($M = 1.42$, $SD = 0.64$) were the lowest.

Participants' interview feedback aligned with the NASA-TLX results, indicating generally low perceived workload during the photo-based conversations. Beyond cognitive workload, participants also commented on interaction comfort and modality quality. Most participants described the glasses-based eye tracker as comfortable (e.g., *P*7 *"felt like a regular pair of glasses,"* and *P*3 noted, *"most of the time I didn't even notice it."*). However, two participants (*P*10, *P*12) anticipated discomfort with prolonged wear; as *P*10 noted, *"I might lose patience if I wore it for a long time."*

6.1.2 Custom User Experience Questionnaire. To assess participants' **user experience**, we asked them to rate the system on a 7-point questionnaire covering three dimensions (14 items in total). Across dimensions, *Overall User Experience* received the highest rating ($M = 6.17$, $SD = 0.75$), followed by *System Effectiveness* ($M = 6.10$, $SD = 0.80$) and *Usability and Accessibility* ($M = 5.81$, $SD = 0.82$). Shapiro–Wilk tests on participant-level dimension means indicated no evidence of non-normality for the Custom User Experience Questionnaire dimensions (all $p > .28$). One-sample t-tests against the scale midpoint (4) further showed that all three dimensions were significantly above neutral: Usability and Accessibility ($t(11) = 8.90$, $p < .001$), System Effectiveness ($t(11) = 11.25$, $p < .001$), and User Experience ($t(11) = 12.62$, $p < .001$). These results suggest that participants perceived the system as usable, effective, and engaging during the reminiscence activities.

6.2 Analysis of Task Performance

During the study, all participants completed the five photo-based conversation blocks. In Section 6.2.1, we summarize how gaze input related to the conversational themes that participants reported. Section 6.2.2 then examines block-wise interaction dynamics using objective measures, including gaze-based engagement and conversational timing.

6.2.1 Gaze-informed prompts align with conversational themes and enable more tailored dialogue. To examine how gaze input relates to conversational themes, we triangulated participants' gaze heatmaps with transcripts of their dialogues with the agent (Fig. 9). For each thematic photo, we first identified a set of salient content themes (e.g., objects/people/places visible in the photo) and then mapped gaze regions of interest (ROIs) to these themes based on the referenced visual elements. This allowed us to compare which photo elements attracted visual attention and which elements became the focus of subsequent conversation.

In the *Childhood* photo, we identified six dominant themes: *People, Furniture, Television, Plants, Heating*, and *Decoration*. Participants whose gaze was concentrated on a small subset of ROIs often produced dialogue that was correspondingly object-specific. For instance, *P*1 primarily attended to *Television* and *Decoration*, and the ensuing dialogue frequently referenced household items

Figure 5: The user study process comprised three phases: (1) the introduction and pre-evaluation, (2) the prototype testing, and (3) semi-structured interviews with the post-evaluation. Each participant participated for approximately 1.5 hours.

Figure 6: In the prototype testing, we selected three representative photos of the live interaction scenarios: (a) Calibration task, (b) Eye tracking recording, and (c) Conversational interaction.

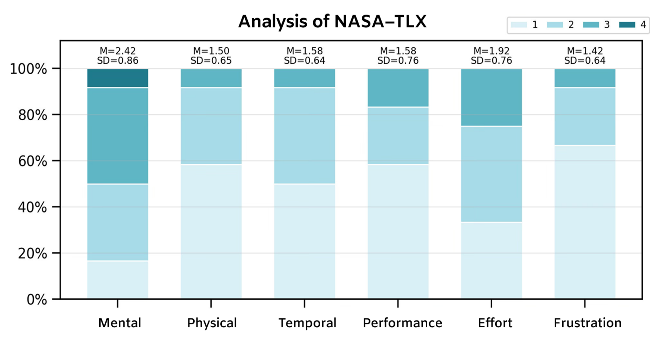


Figure 7: Stacked bar chart of NASA–TLX ratings across six workload dimensions (7-point Likert scale; 1=best, 7=worst; $n = 12$). No responses fell in the high range (scores 5–7).

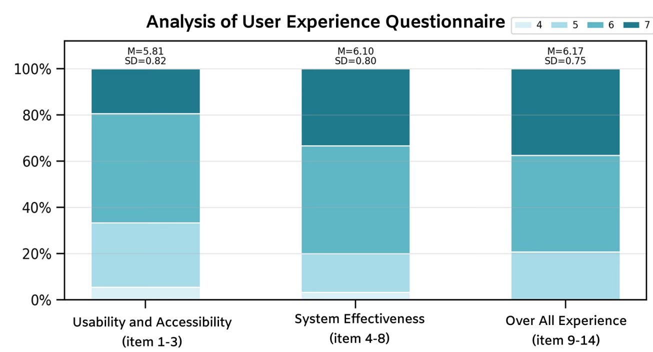


Figure 8: Stacked bar chart of the Custom User Experience Questionnaire ratings (7-point Likert scale; 1=strongly disagree, 7=strongly agree; $n = 12$). No ratings fell in the low range (scores 1–3).

and television-related memories (e.g., *television* and *film* were each mentioned six times in the transcript). Similar photo-element-to-dialogue alignment was observed for *P*3, *P*4, *P*5, *P*9, and *P*10, suggesting that gaze cues can help prioritize content for more tailored follow-up questions.

In contrast, when participants' ROIs were scattered across multiple objects, the system generated more general and less personalized questions. For example, *P*6's ROIs covered nearly all thematic elements, and the agent's questions were correspondingly more general. *P*6's responses contained frequently mentioned but semantically broad words such as *Family* (17 times), *Times* (11 times), and *Child* (9 times). Their low TF–IDF values further indicated less object-specific language.

Participants' Perspectives on Gaze-Driven Personalized Conversations. Most participants expressed satisfaction with the quality of the gaze-driven personalized conversations facilitated by the LLM agent. This was also reflected in ***Q4*** (*The system accurately understands what I say*, $M = 6.00$, $SD = 0.95$) and ***Q6*** (*The system detects my visual interests and directs the conversation accordingly*, $M = 5.58$, $SD = 0.76$). Participants noted that the prototype often steered the dialogue toward content they found personally meaningful. For example, *P*1 explained:

> *"I feel like it truly understands what interests me. It (Agent) starts by asking about TV shows. The old TV brings back so many memories, which makes me keep looking at it."*

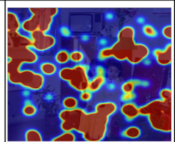
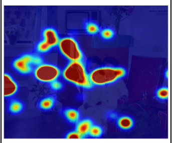
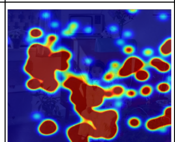
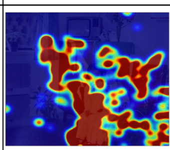
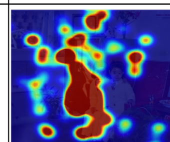
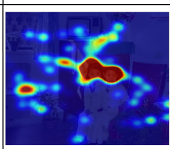
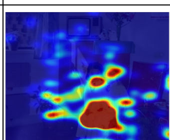
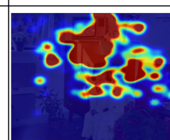
	Heatmap	Question content	Content analysis		Heatmap	Question content	Content analysis
P1		*{ [...] Do you resonate with any of the household items or televisions above these kids, and are there specific details in some of the items or photos that evoke memories of old times? }*	Television 6 0.71% Film 6 0.56% Revolutionary opera 3 0.42% Volleyball 2 0.40% Radio 3 0.30%	P6		*{ [...] Did you ever have items similar to those in the photo, such as the television or furniture? Do the everyday items and decorations in this photo evoke any memories from your childhood or early life? }*	Family 17 0.67% Times 11 0.59% Child 9 0.45% Heating 7 0.38% Television 6 0.35%
Theme	ROI: Television, Decoration	Household Items , Televisions, People	Television, Hobby	Theme	ROI: People, Heating, Decorations, Television	Television, Furniture, Everyday items, Decorations	Family Relationships, Memory Recall
P2		*{ [...] your points of interest are centered around the child and toys above the cabinet. Do these details evoke memories of the time you spent with your child? }*	Toy 10 0.65% Daughter 7 0.58% Child 3 0.35% Television 2 0.25% Past 2 0.15%	P7		*{ [...] Have you ever watched with your family on a similar old-fashioned cupboard, or owned furniture similar to those in the photo? }*	Furniture 8 0.58% Cupboard 8 0.42% Heating 6 0.30% Daughter 5 0.28% Sofa 4 0.24%
Theme	ROI: Decoration, People, Heating, Furniture, Plants	Toys, People	Decoration, People, Television	Theme	ROI: People, Furniture, Plants	Family, Cupboard, Furniture	Furniture, Family Relationships
P3		*{ [...] Do you resonate with any of the household items or television above these kids, and are there specific details in some of the items or photos that evoke memories of old times? }*	Grandma 7 0.45% Family 7 0.40% Classmates 4 0.40% Sister 4 0.35% Child 4 0.35%	P8		*{ [...] Have you grown up in a similar environment or the young girl in the photo that remind you of your own childhood or family life? Or do you remember special moments with your family? }*	Young people 10 0.77% Society 5 0.40% Television 4 0.28% Child 4 0.28% Indifference 4 0.26%
Theme	ROI: People, Decoration, Furniture	Household items, Television, Photos	Family Relationships, Classmates	Theme	ROI: Decoration, People, Television	Young girl, Childhood, Family life	Family Relationships, Childhood
P4		*{ [...] Does seeing the expression on this child's face remind you of a special moment when you were a child? Like watching Television with your family? }*	Television 3 0.34% Furnishing 2 0.20% Furniture 1 0.16% Clothing 1 0.15% Living conditions 1 0.15%	P9		*{ [...] The stuffed toys on the cabinet may have caught your attention. Do you feel that any of the details are related to you? }*	Toy 9 0.68% Daughter 9 0.65% Clothing 5 0.45% Father 2 0.25% Room 1 0.16%
Theme	ROI: People, Heating, Furniture, Plants	Child, Face , Television, Family	Childhood, Family, Hobby	Theme	ROI: Decoration, Television	Toys, Cabinet	Toys, Furniture
P5		*{ [...] Your attention seems to be focused on the little girl, which may be due to her central position [...] Have you recalled a particular moment from your own childhood?}*	Little girl 5 0.49% Father 4 0.39% Occupation 4 0.35% Films 4 0.38% Stamps 3 0.35%	P10		*{ [...] its drawers and cabinet doors likely harboring the cherished possessions of family members. Do you remember having a similar cupboard in your previous home? }*	Cupboard 5 0.57% Television 4 0.47% Little girl 3 0.46% Plant 3 0.40% Memory 2 0.35%
Theme	ROI: People, Clothing, Plants	Little girl, Childhood	Family Relationships, Hobby	Theme	ROI: Furniture, Television, Decoration, People	Drawers, Cupboard, Family members, Wardrobe	Television, Furniture

Figure 9: This figure illustrates example gaze-based heatmaps from 10 participants (out of 12 in total) while viewing *Childhood*-themed photos, together with an analysis of their conversational content. The analysis presents the agent's first-round prompt questions, the frequency counts of the five most common nouns in participants' dialogues with the AI agent, and their corresponding TF–IDF (term frequency–inverse document frequency) values.

Similarly, *P*4 reported that the agent helped surface overlooked details and sustain engagement:

> *"When I used to look at old photos, I often overlooked the details from the past. However, the AI brought up topics that genuinely intrigued me, making me feel more engaged and connected."*

However, two participants (*P*2 and *P*5) raised concerns about whether some questions reliably reflected their gaze. As *P*5 noted, *"I am not sure if all the questions were based on what I saw."* Overall, these responses suggest that gaze-conditioned prompting can make reminiscence conversations feel more personalized, while highlighting the need for more transparent and robust gaze-to-question grounding to build user trust.

6.2.2 Analysis of Gaze Engagement and Conversational Behavior. To assess whether gaze engagement and conversational timing showed systematic changes across the five sequential photo blocks (potential order or time-on-task effects), we analyzed gaze-based engagement (gaze ratio, saccade frequency) and conversational timing (response duration, response latency) across Photo 1–Photo 5 (see Fig. 10); operational definitions and computation details for these metrics are provided in Appendix E. Given the repeated-measures design across Photo 1–Photo 5 and the small sample size, we used nonparametric Friedman tests [94] to assess block-wise differences.

Overall, gaze-based engagement remained stable across Photo 1–Photo 5. **Gaze ratio** was consistently high, ranging from 81.14% to 82.99% (a 1.85 percentage-point span). Mean gaze ratio showed only a slight decrease from Photo 1 (82.81%) to Photo 5 (81.48%). Variability was highest in Photo 5 (SD = 5.28), indicating greater between-participant dispersion in that block rather than a clear monotonic decline over time. Consistent with these descriptive patterns, the Friedman test indicated no significant block effect on gaze ratio, $\chi^2(4) = 6.73$, $p = .151$, Kendall's $W = .14$.

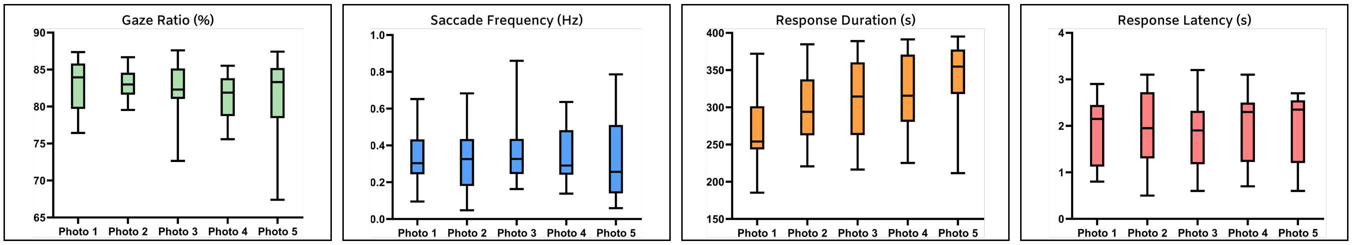


Figure 10: Sequential block effects across five photo blocks (Photo 1–5). Boxplots summarize gaze ratio (%), saccade frequency (Hz), response duration (s), and response latency (s) for $n = 12$.

Saccade frequency was likewise stable across Photo 1–Photo 5 (0.32–0.38 Hz), showing no progressive decrease from Photo 1 to Photo 5 (0.33 to 0.34 Hz). A small descriptive uptick in the middle blocks (Photo 3–4) was observed, but this pattern was not reliable. The Friedman test indicated no significant block effect on **saccade frequency**, $\chi^2(4) = 2.07$, $p = .723$, Kendall's $W = .04$.

In contrast, **response duration** increased across the session, from Photo 1 ($M = 262.98$ s, $SD = 49.24$) to Photo 5 ($M = 341.82$ s, $SD = 51.12$), a net increase of 78.84 s. This increase may reflect participants becoming more comfortable with the activity over time and/or later photos eliciting more elaborate narratives; however, we cannot fully disentangle order effects from photo-specific content in the present design. The Friedman test indicated a significant block effect on response duration, $\chi^2(4) = 25.00$, $p < .001$, Kendall's $W = .52$. **Response latency** remained stable across Photo 1–Photo 5 (means 1.85–2.05 s), with only a small change from Photo 1 to Photo 5 (1.88 to 2.02 s), suggesting no clear slowing in response initiation over time. The Friedman test was not significant for **response latency**, $\chi^2(4) = 3.96$, $p = .411$, Kendall's $W = .08$.

Participants also reported a gradual increase in their willingness to express themselves and sustain attention over the five photo blocks. For example, *P*9 appreciated the chronological sequencing from past to present, describing it as "like screening a series of old photos," and noted that gaze-based interaction felt effortless:

> *"I was satisfied with the ordering of the five photos from past to present. The overall experience was pretty good. Especially, this gaze-based form felt easy for me, because while I was watching, it (Agent) already knew what I was paying attention to"* (*P*9).

In addition, some participants desired greater agency over the conversational flow (*P*1, *P*7, *P*9). As *P*7 remarked, *"I feel that two rounds of dialogue for each photo may not be enough for me"*. These reflections suggest that future versions could offer more user control over parameters such as the number of dialogue turns per photo and the time spent on each image, allowing the system to better accommodate individual pacing and narrative depth.

6.2.3 *Linking log-derived metrics to Subjective Outcomes.*

We explored whether interaction logs related to subjective outcomes by correlating participant-level mean log metrics (gaze ratio, saccade frequency, response latency, response duration) with overall NASA-TLX workload and overall user experience (UX) ($n = 12$). Spearman rank correlations were used and Holm correction [1] was applied within each set of four tests. No associations remained significant after correction (NASA: all $p_{\text{Holm}} \geq .380$; UX: all $p_{\text{Holm}} \geq .492$). The strongest uncorrected effects were response latency with NASA ($\rho = -.504$, $p = .095$, $p_{\text{Holm}} = .380$) and gaze ratio with UX ($\rho = -.470$, $p = .123$, $p_{\text{Holm}} = .492$).

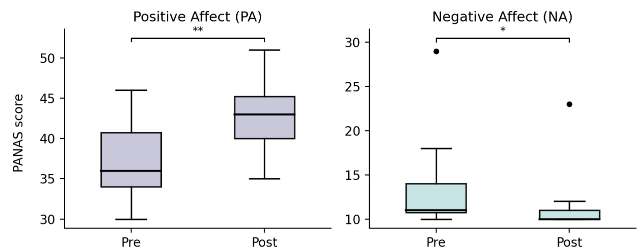


Figure 11: PANAS pre–post score results ($n = 12$): Positive affect (PA) increased and negative affect (NA) decreased. Statistical significance is indicated as $^{*}p < .05$, $^{**}p < .01$.

6.3 Emotional and Reflective Benefits

6.3.1 *Enhancing emotional well-being through positive memory recall.*

To examine whether interacting with *Eye2Recall* was associated with short-term affective changes, we assessed participants' pre–post affect using PANAS. Overall, participants reported higher positive affect and lower negative affect immediately after the brief session. We assessed affect pre–post using PANAS (see Fig. 11). **Positive Affect (PA)** increased from $M_{\text{pre}} = 37.42$ ($SD = 5.25$) to $M_{\text{post}} = 42.75$ ($SD = 4.39$); a paired-samples t-test indicated a significant improvement, $t(11) = 3.17$, $p = .009$, with a large within-subject effect size ($d_z = 0.92$). **Negative Affect (NA)** decreased from $M_{\text{pre}} = 13.42$ ($SD = 5.43$) to $M_{\text{post}} = 11.50$ ($SD = 3.68$). Given the small sample size, we additionally ran Wilcoxon signed-rank tests as a robustness check; results were directionally consistent (PA: $p = .012$; NA: $p = .022$).

Participants described their experience with positive descriptions such as *"cozy," "warm,"* (*P*2, *P*6) and *"touching,"* (*P*4, *P*9), likening it to *"a dialogue between time and space"* (*P*5). Participants felt that the system provides meaningful memories and emotional experiences (***Q7****, The system provides meaningful memories and emotional experiences.* $M = 6.17$, $SD = 0.58$). After using the system, they felt happy, and their psychological well-being was enhanced (***Q8****, After using the system, I feel happy and my psychological well-being is enhanced.*

$M = 6.17$, $SD = 0.80$). $P4$ added *"it gave me much encouragement and was willing to listen to my story, it makes me feel happy"*. $P6$ commented that:

> *"In everyday life, my children work in other cities, so I rarely have someone who truly listens to me. It felt wonderful to have someone willing to hear me out. When I told the story of how my mother struggled to support our whole family when she was young, the agent gave her heartfelt recognition, and I was deeply moved."*

6.3.2 Deepening reflection through historical context provided by the LLM agent. By presenting brief, photo-linked historical prompts and then asking specific questions, the system encouraged reflective remembrance, helping participants to connect personal experiences to wider social change, organise their stories and express their meanings rather than simply recalling facts. Participants shifted their focus to broader social changes, such as household technologies, festive practices and neighbourhood life, when interacting with the LLM agent. Participants connected personal episodes to broader social change and articulated meanings beyond factual recall. P5 called these reflections *"priceless treasures,"* while P6 said:

> *"After talking with the agent, I rediscovered the atmosphere of Spring Festival in my childhood, even though life was difficult at that time,"*

noting that brief introduction about regional folk customs felt *"intimate."* $P11$ similarly remarked that the experience made previously overlooked parts of the past more salient and memorable.

6.4 Suggestions for System Improvement

6.4.1 Enhancing emotional responsiveness in reminiscence dialogue. The agent's encouraging style during dialogue appeared to support affective engagement and reflection. Several participants ($P1$, $P6$, $P8$, $P11$) suggested that improving emotion detection and response would make the experience more engaging. As $P11$ put it,

> *"I hoped the agent could vary its prosody and offer more personalized reactions based on user feedback; for example, when I recounting a wedding story, the agent's tone should sound warm and celebratory to match the moment."*

Two participants ($P6$ and $P8$) became tearful and choked up during the conversational phase. $P8$ described the experience as immersive and emotionally real, explaining that *"I cried mainly because I was moved, and I felt encouraged."* $P8$ reflected that recalling time with parents *"made me feel stronger and happier."* P6 also noted,

> *"If the system could pick up on my emotions, for example through my tone or expressions, and respond to that, I would feel more understood."*

Both participants suggested that greater emotional responsiveness could further improve the experience. In practice this means recognizing cues of sadness or joy, acknowledging the user's feelings in plain language, adjusting the pace and length of turns, offering an optional pause, and proposing supportive follow-ups such as a gentler question or a transition to a neutral topic when needed.

6.4.2 Documenting and sharing recalled memories to foster intergenerational communication. Most participants (7 of 12) expressed a desire to record and selectively share their reminiscence outputs with younger family members, aiming to support intergenerational connection and mutual understanding. As P1 said,

> *"I hope my children can hear the sound of my memories when they grow up; it's truly valuable for our family."*

$P10$ added that, if sustained over time, the experience could become *"a repository of family memories."* We interpret this as a design implication for optional, consent-based archiving (e.g., audio clips, transcripts) with fine-grained controls over what is shared, with whom, and the ability to retract sharing at any time.

7 Discussion

Our findings suggest that *Eye2Recall* can support older adults' photo-based reminiscence through customized, fluent interactions and perceived well-being benefits. We next discuss how gaze cues enrich reminiscence, address ethical, privacy, and safety considerations, present four design implications, and conclude with limitations and future directions.

7.1 Leveraging Gaze Cues with LLMs to Enrich Photo-based Reminiscence

Prior work has shown that eye movements are closely involved in the retrieval of autobiographical memories [21, 69, 89]. However, despite growing interest in attention-aware interfaces, the integration of eye-tracking signals with LLM-driven conversational prompting to support older adults' photo-based reminiscence remains underexplored [45].

Our work addresses this gap by combining gaze-driven initiative signals with an LLM facilitator to enable low-effort, mixed-initiative reminiscence conversations. Specifically, we examine how gaze-based cues can steer an LLM-powered conversational agent during photo-based reminiscence. Building on prior work showing that gaze supports joint attention and reveals what visual elements are salient to an interlocutor, we use gaze-defined ROIs to guide prompt selection and constrain the focus of follow-up questions for older adults.

In our study, participants generally articulated concrete personal memories, and most (10 of 12) reported that the agent's topic initiation aligned with what they found interesting in the photo (see Section 6.2.1). We interpret this as suggestive evidence that gaze-conditioned prompts may reduce the effort needed to find a conversational entry point and help the agent stay on user-salient elements, rather than as a definitive claim of superiority over alternative inputs.

At the same time, our implementation was intentionally conservative: gaze was used only to seed the first gaze-conditioned turn (Prompt 2) and to focus the initial set of follow-up questions, rather than as a continuous signal across turns. This choice limited opportunities to adapt mid-conversation (e.g., re-focusing after topic drift or adjusting to emergent interests). Future research should investigate how to integrate eye-tracking continuously throughout extended conversations, including when to update gaze-conditioned focus and how to balance gaze cues with conversational context, to enable more dynamic, responsive, and context-aware conversational systems.

7.2 Ethical, Privacy, and Safety Considerations

7.2.1 Safety-by-Design and Human Support Pathways. Safety-by-design is equally critical given the emotionally evocative nature of reminiscence. We suggest setting clear boundaries on the agent's role (e.g., it does not provide medical or legal advice), providing plain-language explanations of system capabilities and limitations, and offering always-available controls to **pause, skip, or redirect** sensitive topics. Because some participants became tearful during conversations, deployments should also include pathways to human support (e.g., caregiver escalation or on-call staff in institutional settings) when distress is observed or reported. Finally, these safeguards should be evaluated with diverse older-adult populations and in ecologically valid contexts (including home use), balancing potential benefits with privacy, autonomy, and cultural expectations.

7.2.2 Reminiscence Material Selection. Prior reminiscence research in HCI field has used a wide range of *external cues* to stimulate memory and support recalling the past, including music [35], cultural heritage artefacts [58], and social media content [62]. Old photographs are also widely used as cues; however, much of this work relies on *personal* photo collections [5, 55]. Zhongyue et al. noted that integrating personal photos into interactive reminiscence experiences can introduce practical and socio-technical challenges, such as quality issues, long-term preservation and management of family archives, and photo annotation [93].

Against this backdrop, our study intentionally used *non-personal*, era-typical historical photos as external cues [8, 11, 91]. This design choice enabled culturally shared prompts that many participants could relate to without requiring access to private family archives, while also standardizing session content across participants. Despite their practical advantages, comparatively few systems have examined non-personal old photographs as reminiscence cues. Our findings suggest that such photos can serve as effective, low-barrier prompts for initiating and sustaining reminiscence conversations, while sidestepping some of the privacy and access constraints associated with personal albums.

7.2.3 Privacy and Data Governance. Privacy and data governance in gaze- and voice-mediated reminiscence require more than simply storing data locally. In our study, we followed a data-minimization and purpose-limitation approach, collecting only what was necessary to support the reminiscence session. We further recommend that future deployments incorporate explicit access controls, clear retention limits with secure deletion, and lightweight audit trails that document when data are accessed or exported, to support transparency and accountability.

7.3 Design Implications

We proposed four design implications (DIs) derived from the key challenges participants faced and their expectations for improvement.

7.3.1 Emotion-aware pacing and tone as an optional scaffold. Rather than full real-time sentiment classification, we recommend lightweight, privacy-preserving adaptation that leverages gaze-derived signals of interest (e.g., sustained dwell on regions of interest, refixations, or repeated attention shifts). Based on these cues, the backend can proactively offer timely prompts (e.g., "Would you like to continue with this photo or move on?") and adjust pacing, acknowledgements, and tone without inferring fine-grained emotions.

Crucially, control should remain with the user: older adults should be able to use gaze to select on-screen UI options to **skip** or **switch** photos when content feels uninteresting or uncomfortable. Providing adjustable sensitivity settings (e.g., how quickly the system prompts a change) can further align the interaction with individual preferences while mitigating misclassification risk. Future systems could provide optional, user-controlled affect-aware pacing and tone adjustments, alongside customizable voice preferences (e.g., selecting a male or female voice) to better match individual comfort and listening habits.

7.3.2 Sharing and controlling consent-based reminiscence content and digital legacies. Participants expressed a desire to preserve memories related to photos that reflect their life journey, and build a digital legacy (see Section 6.4.2). They hope to create a collection of personal digital assets to share with their children and future generations. Digital legacy materials hold significant values and meanings that can be passed down through generations [20]. It is increasingly important for researchers to consider their systems' potential role in preserving and transmitting digital legacies [53]. The system should allow users to easily archive recalled memories, including text, photos, and audio, into a personal digital archive, creating a meaningful and comprehensive digital legacy that preserves life stories for future generations.

7.3.3 Robust multimodal pipeline for stable, fluent interaction. Our study revealed that the system struggled to recognize Chinese text in images due to optical character recognition (OCR) limitations, leading to misinterpretations. To address this, integrating more advanced OCR models tailored for multilingual recognition can enhance accuracy. Moreover, conversational breakdowns, such as abrupt interruptions and unnatural pauses, suggest the need for hybrid AI models that blend rule-based approaches with machine learning for more stable and responsive dialogues. Additionally, reducing latency in text-to-speech conversion is critical, particularly for users interacting in a second language, to ensure fluid and natural communication.

7.3.4 Accessibility by device and modality diversity. Two participants reported discomfort with head-mounted eye tracking, and some wore glasses (see Section 6.1). To reduce device burden and broaden accessibility, future systems should offer alternative setups, such as remote (screen-based) eye tracking, camera-based gaze proxies, and gaze-free modes that rely on voice with simple pointing or on-screen highlighting. These options should follow universal design principles (e.g., larger fonts, high contrast, adjustable volume and speaking rate) and allow users to switch modalities mid-session as comfort and needs change.

7.4 Limitations and Future Work

Our study has several limitations. First, the study relied on a small sample of 12 Chinese older adults, which limits the generalizability of our findings. Individual differences in education, digital literacy, cognitive and physical abilities, and cultural background may affect how older adults respond to gaze-driven prompts and recall memories, suggesting the need for future studies with more diverse participants and adaptive support.

Second, our study was designed as an exploratory investigation of gaze-driven prompting to support reminiscence with old photos, and therefore did not include a comparison condition. As a result, we cannot isolate the specific contribution of gaze-driven prompting from the general benefits of reminiscence activities. Future studies should incorporate controlled baselines that vary only the triggering method, such as alternative non-eye-gaze triggers (e.g., pointing, tapping) or a no-trigger baseline, while keeping all other conditions identical.

Third, our study was short-term and conducted in a laboratory setting, limiting conclusions about the durability of effects and everyday adoption. Future work should evaluate *Eye2Recall* in naturalistic, home-based deployments with repeated follow-ups and validated measures of cognition and well-being, ideally analyzed using mixed-effects models with reported effect sizes and power analyses. Moreover, our current setup used a large display and standardized, era-typical photos, which may differ from real-world use with personal photo collections and various devices (e.g., tablets or smart TVs) where lighting, familiarity, distractions, and privacy vary. Future studies should account for these real-world factors and strengthen safeguards for sensitive memories and bystander privacy.

8 Conclusion

We introduce *Eye2Recall*, a novel gaze-driven, LLM-supported system designed to facilitate reminiscence with old photos for older adults. To inform its design, we conducted interviews with four domain experts, whose insights guided the creation of an accessible and engaging prototype grounded in user-centered design principles, accessibility, and empowerment. The resulting system combines eye-tracking technology with LLM-based conversation, addressing a research gap in mixed-initiative interaction. Evaluation with older adult participants demonstrated positive user experiences, and their feedback will inform refinements to enhance consistency and usability in future studies.

Acknowledgments

This work was supported by the Computational Media and Arts Lab[6] and the Neural Encoding & Decoding Lab[7]. We thank Dr. Kefei Liu, Assistant Professor in Bioscience and Biomedical Engineering at HKUST (Guangzhou), for his support throughout the research. We also thank Dr. Rongrong Chen, Assistant Professor at Beijing Normal University–Hong Kong Baptist University (BNBU), for her valuable guidance during the research process.

[6]https://cma.hkust-gz.edu.cn/about-cma/

[7]https://kefeilab.hkust-gz.edu.cn/Neural-Encoding-Decoding-Lab

References

[1] Hervé Abdi. 2010. Holm's sequential Bonferroni procedure. *Encyclopedia of research design* 1, 8 (2010), 1–8.

[2] M Argyle and M Cook. 1976. Gaze and mutual gaze.- Cambridge University Press.

[3] Arlene J Astell, Maggie P Ellis, Norman Alm, Richard Dye, and Gary Gowans. 2010. Stimulating people with dementia to reminisce using personal and generic photographs. *International Journal of Computers in Healthcare* 1, 2 (2010), 177–198.

[4] Benett Axtell and Cosmin Munteanu. 2018. Frame of Mind: Using Storytelling for Speech-Based Clustering of Family Pictures. In *Companion Proceedings of the 23rd International Conference on Intelligent User Interfaces* (Tokyo, Japan) *(IUI '18 Companion)*. Association for Computing Machinery, New York, NY, USA, Article 22, 2 pages. doi:10.1145/3180308.3180330

[5] Benett Axtell and Cosmin Munteanu. 2019. PhotoFlow in Action: Picture-Mediated Reminiscence Supporting Family Socio-Connectivity. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–4. doi:10.1145/3290607.3313272

[6] Long Bai, Xiangfei Liu, and Jiacan Su. 2023. ChatGPT: The Cognitive Effects on Learning and Memory. *Brain-X* 2, Article 30 (Sept. 2023). doi:10.1002/brx2.30

[7] Alistair Baron, Paul Rayson, Dawn Archer, et al. 2009. Word frequency and key word statistics in historical corpus linguistics. *Anglistik: International Journal of English Studies* 20, 1 (2009), 41–67.

[8] Michael P. Bender, Andrew Norris, and Paulette Bauckham. 1998. *The Therapeutic Purposes of Reminiscence.* Sage Publications, London.

[9] James E Birren and Donna E Deutchman. 1991. *Guiding autobiography groups for older adults: Exploring the fabric of life.* JHU Press, Baltimore, MD.

[10] Ernst Bohlmeijer, Marte Roemer, Pim Cuijpers, and Filip Smit. 2007. The Effects of Reminiscence on Psychological Well-Being in Older Adults: A Meta-Analysis. *Aging Ment. Health* 11, 3 (May 2007), 291–300. doi:10.1080/13607860600963547

[11] Ernst Bohlmeijer, Marije Valenkamp, Gerben Westerhof, Filip Smit, and Pim Cuijpers. 2005. Creative reminiscence as an early intervention for depression: Results of a pilot project. *Aging & Mental Health* 9, 4 (2005), 302–304.

[12] M Christine Boyer. 1994. *The city of collective memory: its historical imagery and architectural entertainments.* MIT Press.

[13] John Brooke et al. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7.

[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] Sharon Lynn Chu, Brittany Garcia, Taylor Quance, Lisa Geraci, Steven Woltering, and Francis Quek. 2016. Understanding storytelling as a design framework for cognitive support technologies for older adults. In *Proceedings of the International Symposium on Interactive Technology and Ageing Populations.* 24–33.

[16] John R Crawford and Julie D Henry. 2004. The positive and negative affect schedule (PANAS): construct validity, measurement properties and normative data in a large non-clinical sample. *British Journal of Clinical Psychology* 43, 3 (2004), 245–265.

[17] Florian Daniel, Jin Yu, Boualem Benatallah, Fabio Casati, Maristella Matera, and Regis Saint-Paul. 2007. Understanding UI Integration: A Survey of Problems, Technologies, and Opportunities. *IEEE Internet Computing* 11, 3 (2007), 59–66. doi:10.1109/MIC.2007.74

[18] Tanya E Davison, Kanvar Nayer, Selby Coxon, Arthur de Bono, Barbara Eppingstall, Yun-Hee Jeon, Eva S van der Ploeg, and Daniel W O'Connor. 2016. A personalized multimedia device to treat agitated behavior and improve mood in people with dementia: A pilot study. *Geriatric Nursing* 37, 1 (2016), 25–29.

[19] Lina Dib, Daniela Petrelli, and Steve Whittaker. 2010. Sonic souvenirs: exploring the paradoxes of recorded sound for family remembering. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work* (Savannah, Georgia, USA) *(CSCW '10)*. Association for Computing Machinery, New York, NY, USA, 391–400. doi:10.1145/1718918.1718985

[20] Dylan Thomas Doyle and Jed R. Brubaker. 2023. Digital Legacy: A Systematic Literature Review. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW2, Article 268 (Oct. 2023), 26 pages. doi:10.1145/3610059

[21] Mohamad El Haj. 2024. When you look at your past: Eye movement during autobiographical retrieval. *Consciousness and Cognition* 118 (2024), 103652.

[22] Ajoy S. Fernandes, Immo Schütz, T. Scott Murdison, and Michael J. Proulx. 2025. Gaze Inputs for Targeting: The Eyes Have It, Not With a Cursor. *International Journal of Human–Computer Interaction* 41, 19 (2025), 12251–12269. doi:10.1080/10447318.2025.2453966 arXiv:https://doi.org/10.1080/10447318.2025.2453966

[23] Andrea Ferrario, Burcu Demiray, Kristina Yordanova, Minxia Luo, and Mike Martin. 2020. Social Reminiscence in Older Adults' Everyday Conversations: Automated Detection Using Natural Language Processing and Machine Learning. *J. Med. Internet Res.* 22, 9, Article e19133 (Sept. 2020). doi:10.2196/19133

[24] Joseph M Fitzgerald. 1988. Vivid memories and the reminiscence phenomenon: the role of a self narrative. *Human Development* 31, 5 (1988), 261–273.

[25] Julie Fleury, Constantine Sedikides, Tim Wildschut, David W. Coon, and Pauline Komnenich. 2022. Feeling Safe and Nostalgia in Healthy Aging. *Frontiers in Psychology* Volume 13 - 2022 (2022). doi:10.3389/fpsyg.2022.843051

[26] Kim Flottemesch. 2013. Learning through narratives: The impact of digital storytelling on intergenerational relationships. *Academy of Educational Leadership Journal* 17, 3 (2013), 53.

[27] James J Gross. 2015. Emotion regulation: Current status and future prospects. *Psychological Inquiry* 26, 1 (2015), 1–26.

[28] Lei Han, Yu Zhou, Qiongyan Chen, and David Yip. 2024. Memory Remedy: An AI-Enhanced Interactive Story Exploring Human-Robot Interaction and Companionship. In *Proceedings of the 17th International Symposium on Visual Information Communication and Interaction (VINCI '24)*. Association for Computing Machinery, New York, NY, USA, Article 16, 5 pages. doi:10.1145/3678698.3687186

[29] Celia B. Harris, Penny Van Bergen, Paul A. Strutt, Gabrielle K. Picard, Sophia A. Harris, Ruth Brookman, and Karn Nelson. 2022. Teaching Elaborative Reminiscing to Support Autobiographical Memory and Relationships in Residential and Community Aged Care Services. *Brain Sciences* 12, 3, Article 374 (March 2022). doi:10.3390/brainsci12030374

[30] Sandra G Hart. 2006. NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 50. Sage publications Sage CA: Los Angeles, CA, 904–908.

[31] Kasper Hornbæk, Aske Mottelson, Jarrod Knibbe, and Daniel Vogel. 2019. What Do We Mean by "Interaction"? An Analysis of 35 Years of CHI. *ACM Trans. Comput.-Hum. Interact.* 26, 4, Article 27 (July 2019), 30 pages. doi:10.1145/3325285

[32] Minlie Huang, Xiaoyan Zhu, and Jianfeng Gao. 2020. Challenges in Building Intelligent Open-domain Dialog Systems. *ACM Trans. Inf. Syst.* 38, 3, Article 21 (apr 2020), 32 pages.

[33] Robert J. K. Jacob. 1995. *Eye tracking in advanced interface design.* Oxford University Press, Inc., USA, 258–288.

[34] Keisha Jayaratne. 2016. The Memory Tree: Using Sound to Support Reminiscence. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI EA '16)*. Association for Computing Machinery, New York, NY, USA, 116–121. doi:10.1145/2851581.2890384

[35] Yucheng Jin, Wanling Cai, Li Chen, Yizhe Zhang, Gavin Doherty, and Tonglin Jiang. 2024. Exploring the Design of Generative AI in Supporting Music-based Reminiscence for Older Adults. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 1012, 17 pages. doi:10.1145/3613904.3642800

[36] Callie Y. Kim, Christine P. Lee, and Bilge Mutlu. 2024. Understanding Large-Language Model (LLM)-powered Human-Robot Interaction. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction* (Boulder, CO, USA) *(HRI '24)*. Association for Computing Machinery, New York, NY, USA, 371–380. doi:10.1145/3610977.3634966

[37] Jeong Kim and John Zimmerman. 2006. Cherish: smart digital photo frames for sharing social narratives at home. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems* (Montréal, Québec, Canada) *(CHI EA '06)*. Association for Computing Machinery, New York, NY, USA, 953–958. doi:10.1145/1125451.1125635

[38] Chris L. Kleinke. 1986. Gaze and Eye Contact: A Research Review. *Psychological Bulletin* 100, 1 (1986), 78–100. doi:10.1037/0033-2909.100.1.78

[39] Liliana Laranjo, Adam G. Dunn, Huong Ly Tong, Ahmet Baki Kocaballi, Jessica Chen, Rabia Bashir, Didi Surian, Blanca Gallego, Farah Magrabi, Annie Y. S. Lau, and Enrico Coiera. 2018. Conversational Agents in Healthcare: A Systematic Review. *J. Am. Med. Inform. Assoc.* 25, 9 (Sept. 2018), 1248–1258. doi:10.1093/jamia/ocy072

[40] Amanda Lazar, Hilaire Thompson, and George Demiris. 2014. A systematic review of the use of technology for reminiscence therapy. *Health education & behavior* 41, 1_suppl (Oct. 2014), 51S–61S. doi:10.1177/1090198114537067

[41] Minha Lee, Sander Ackermans, Nena van As, Hanwen Chang, Enzo Lucas, and Wijnand IJsselsteijn. 2019. Caring for Vincent: A Chatbot for Self-Compassion. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3290605.3300932

[42] Charles N. Lewis. 1971. Reminiscing and Self-Concept in Old Age. *Journal of Gerontology* 26, 2 (apr 1971), 240–243. doi:10.1093/geronj/26.2.240

[43] Hanzhi Li, Jianping Jia, and Zhiqiang Yang. 2016. Mini-mental state examination in elderly Chinese: a population-based normative study. *Journal of Alzheimer's Disease* 53, 2 (2016), 487–496.

[44] Minglan Li, Yipeng Yu, Xu Liu, Junqing Wu, Qiandong Wang, and Yueqin Hu. 2025. Beyond algorithms: Utilizing multi-modal emotional and behavioral cues as novel predictors of short-video consumption. *Computers in Human Behavior Reports* 20 (2025), 100805. doi:10.1016/j.chbr.2025.100805

[45] Ruohao Li, Jiawei Li, Jia Sun, Zhiqing Wu, Zisu Li, Ziyan Wang, Ge Lin, and Mingming Fan. 2025. RemVerse: Supporting Reminiscence Activities for Older Adults through AI-Assisted Virtual Reality. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 9, 3, Article 103 (Sept. 2025), 25 pages. doi:10.1145/3749505

[46] Zisu Li, Li Feng, Chen Liang, Yuru Huang, and Mingming Fan. 2023. Exploring the Opportunities of AR for Enriching Storytelling with Family Photos between Grandparents and Grandchildren. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 7, 3, Article 108 (Sept. 2023), 26 pages. doi:10.1145/3610903

[47] Guangzhou Digital Library. 2023. Guangzhou Digital Library. https://www.gzlib.org.cn. Accessed: May 20, 2023.

[48] Haining Liu, Haihong Liu, Feng Li, Buxin Han, and Cuili Wang. 2021. Effect of Cognitive Control on Attentional Processing of Emotional Information Among Older Adults: Evidence From an Eye-Tracking Study. *Front. Aging Neurosci.* 13, Article 644379 (April 2021). doi:10.3389/fnagi.2021.644379

[49] Ángela López-Cardona, Carlos Segura, Alexandros Karatzoglou, Sergi Abadal, and Ioannis Arapakis. 2025. Seeing Eye to AI: Human Alignment via Gaze-Based Response Rewards for Large Language Models. In *International Conference on Representation Learning*, Y. Yue, A. Garg, N. Peng, F. Sha, and R. Yu (Eds.), Vol. 2025. 26111–26135. https://proceedings.iclr.cc/paper_files/paper/2025/file/4165875c8140617b8c165e18c342ccbb-Paper-Conference.pdf

[50] Lara de Sá Neves Loureiro, Maria das Graças Melo Fernandes, Sueli Marques, Maria Miriam Lima da Nobrega, and Rosalina A Partezani Rodrigues. 2013. Burden in family caregivers of the elderly: prevalence and association with characteristics of the elderly and the caregivers. *Revista Da Escola De Enfermagem Da USP* 47 (2013), 1129–1136.

[51] Dongye Lyu, Luis Manas-Viniegra, and Ziyuan Xu. 2025. Visual attention differences toward football stadium's naming rights: an eye tracking study. *Asia Pacific Journal of Marketing and Logistics* 37, 1 (2025), 189–209.

[52] Anna Madoglou, Theofilos Gkinopoulos, Panagiotis Xanthopoulos, and Dimitrios Kalamaras. 2017. Representations of autobiographical nostalgic memories: Generational effect, gender, nostalgia proneness and communication of nostalgic experiences. *Journal of Integrated Social Sciences* 7, 1 (2017), 60–88.

[53] Michael Massimi, William Odom, Richard Banks, and David Kirk. 2011. Matters of life and death: locating the end of life in lifespan-oriented hci research. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. Association for Computing Machinery, New York, NY, USA, 987–996. doi:10.1145/1978942.1979090

[54] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. 2019. Reliability and Inter-rater Reliability in Qualitative Research: Norms and Guidelines for CSCW and HCI Practice. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 72 (Nov. 2019), 23 pages. doi:10.1145/3359174

[55] 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) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3290605.3300665

[56] Walaa Medhat, Ahmed Hassan, and Hoda Korashy. 2014. Sentiment analysis algorithms and applications: A survey. *Ain Shams engineering journal* 5, 4 (2014), 1093–1113.

[57] Chelsea Myers, Anushay Furqan, Jessica Nebolsky, Karina Caro, and Jichen Zhu. 2018. Patterns for How Users Overcome Obstacles in Voice User Interfaces. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–7. doi:10.1145/3173574.3173580

[58] Àngela Nebot, Sara Domènech, Natália Albino-Pires, Francisco Mugica, Anass Benali, Xènia Porta, Oriol Nebot, and Pedro M Santos. 2022. LONG-REMI: an AI-based technological application to promote healthy mental longevity grounded in reminiscence therapy. *International Journal of Environmental Research and Public Health* 19, 10 (2022), 5997.

[59] National Library of China. 2023. National Library of China. https://www.nlc.cn/web/index.shtml. Accessed: May 21, 2023.

[60] Viral Parekh, Pin Sym Foong, Shengdong Zhao, and Ramanathan Subramanian. 2018. AVEID: Automatic Video System for Measuring Engagement In Dementia. In *Proceedings of the 23rd International Conference on Intelligent User Interfaces* (Tokyo, Japan) *(IUI '18)*. Association for Computing Machinery, New York, NY, USA, 409–413. doi:10.1145/3172944.3173010

[61] Katarina Pavic, Ali Oker, Mohamed Chetouani, and Laurence Chaby. 2021. Age-related Changes in Gaze Behaviour During Social Interaction: An Eye-Tracking Study with an Embodied Conversational Agent. *Q. J. Exp. Psychol.* 74, 6 (June 2021), 1128–1139. doi:10.1177/1747021820982165

[62] 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) *(CHI '10)*. Association for Computing Machinery, New York, NY, USA, 2027–2036. doi:10.1145/1753326.1753635

[63] Daniela Petrelli, Steve Whittaker, and Jens Brockmeier. 2008. AutoTopography: what can physical mementos tell us about digital memories?. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) *(CHI '08)*. Association for Computing Machinery, New York, NY, USA, 53–62. doi:10.1145/1357054.1357065

[64] Martin Pinquart and Simon Forstmeier. 2012. Effects of reminiscence interventions on psychosocial outcomes: A meta-analysis. *Aging & mental health* 16, 5 (2012), 541–558.

[65] Aditya kumar Purohit, Aditya Upadhyaya, and Adrian Holzer. 2023. ChatGPT in Healthcare: Exploring AI Chatbot for Spontaneous Word Retrieval in Aphasia. In *Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing* (Minneapolis, MN, USA) *(CSCW '23 Companion)*. Association for Computing Machinery, New York, NY, USA, 1–5. doi:10.1145/3584931.3606993

[66] Pernilla Qvarfordt. 2017. *Gaze-informed multimodal interaction.* Association for Computing Machinery and Morgan & Claypool, 365–402. https://doi.org/10.1145/3015783.3015794

[67] Michael Romaniuk and Jean Gasen Romaniuk. 1981. Looking back: An analysis of reminiscence functions and triggers. *Experimental aging research* 7, 4 (1981), 477–489. doi:10.1080/03610738108259826

[68] Hyeyoung Ryu, Soyeon Kim, Dain Kim, Sooan Han, Keeheon Lee, and Younah Kang. 2020. Simple and Steady Interactions Win the Healthy Mentality: Designing a Chatbot Service for the Elderly. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW2, Article 152 (Oct. 2020), 25 pages. doi:10.1145/3415223

[69] Ali Salamatian, Amirhossein Abaskohi, Wan-Cyuan Fan, Mir Rayat Imtiaz Hossain, Leonid Sigal, and Giuseppe Carenini. 2025. ChartGaze: Enhancing Chart Understanding in LVLMs with Eye-Tracking Guided Attention Refinement. arXiv:2509.13282 [cs.CL] https://arxiv.org/abs/2509.13282

[70] Jesús Sánchez Cuadrado, Sara Pérez-Soler, Esther Guerra, and Juan De Lara. 2024. Automating the Development of Task-oriented LLM-based Chatbots. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces* (Luxembourg, Luxembourg) *(CUI '24)*. Association for Computing Machinery, New York, NY, USA, Article 11, 10 pages. doi:10.1145/3640794.3665538

[71] Stefan C Schmukle, Boris Egloff, and Lawrence R Burns. 2002. The relationship between positive and negative affect in the Positive and Negative Affect Schedule. *Journal of Research in Personality* 36, 5 (2002), 463–475.

[72] Constantine Sedikides, Tim Wildschut, Clay Routledge, Jamie Arndt, Erica G Hepper, and Xinyue Zhou. 2015. To nostalgize: Mixing memory with affect and desire. In *Advances in Experimental Social Psychology*. Vol. 51. Elsevier, 189–273.

[73] Thibaut Septon, Théo Leclercq, and Bruno Dumas. 2025. Leveraging Pronoun Disambiguation in Multimodal Interaction for Contextual Understanding of Voice Assistant Queries. In *Companion Proceedings of the 30th International Conference on Intelligent User Interfaces (IUI '25 Companion)*. Association for Computing Machinery, New York, NY, USA, 155–158. doi:10.1145/3708557.3716362

[74] Azmeh Shahid, Kate Wilkinson, Shai Marcu, and Colin M Shapiro. 2012. Stanford sleepiness scale (SSS). *STOP, THAT and One Hundred Other Sleep Scales* (2012), 369–370.

[75] Lu Shaoming. 2020. The exploration of memory place protection from local perspective. *China Ancient City* 7 (2020), 4–9.

[76] Beijing Silvermine. 2023. Beijing Silvermine. https://www.beijingsilvermine.com. Accessed: May 20, 2023.

[77] Brodrick Stigall, Jenny Waycott, Steven Baker, and Kelly Caine. 2019. Older adults' perception and use of voice user interfaces: a preliminary review of the computing literature. In *Proceedings of the 31st Australian Conference on Human-Computer-Interaction*. 423–427.

[78] Annalisa Szymanski, Noah Ziems, Heather A. Eicher-Miller, Toby Jia-Jun Li, Meng Jiang, and Ronald A. Metoyer. 2025. Limitations of the LLM-as-a-Judge Approach for Evaluating LLM Outputs in Expert Knowledge Tasks. In *Proceedings of the 30th International Conference on Intelligent User Interfaces (IUI '25)*. Association for Computing Machinery, New York, NY, USA, 952–966. doi:10.1145/3708359.3712091

[79] Josephine Rose Orejana Tan, Petra Boersma, Teake P Ettema, Caroline HM Planting, Soraya Clark, Robbert JJ Gobbens, and Rose-Marie Dröes. 2023. The effects of psychosocial interventions using generic photos on social interaction, mood and quality of life of persons with dementia: a systematic review. *BMC geriatrics* 23, 1 (2023), 560.

[80] Seiki Tokunaga, Kazuhiro Tamura, and Mihoko Otake-Matsuura. 2021. A Dialogue-Based System with Photo and Storytelling for Older Adults: Toward Daily Cognitive Training. *Frontiers in Robotics and AI* Volume 8 - 2021 (2021). doi:10.3389/frobt.2021.644964

[81] Abel Toledano-González, Dulce Romero-Ayuso, Dolores Fernández-Pérez, Marta Nieto, Jorge Javier Ricarte, Beatriz Navarro-Bravo, Laura Ros, and José Miguel Latorre. 2023. Effects of the use of autobiographical photographs on emotional induction in older adults: a systematic review. *Psychological Research* 87, 4 (2023), 988–1011.

[82] Karolina Trajkovska, Matjaž Kljun, and Klen Čopič Pucihar. 2024. Gaze2AOI: Open Source Deep-learning Based System for Automatic Area of Interest Annotation with Eye Tracking Data. arXiv:2411.13346 [cs.SE] https://arxiv.org/abs/2411.13346

[83] E Tulving. 1983. Elements of episodic memory.

[84] Moshe Unger, Alexander Tuzhilin, and Michel Wedel. 2025. STARE: Predicting Decision Making Based on Spatio-Temporal Eye Movements. arXiv:2508.04148 [cs.HC] https://arxiv.org/abs/2508.04148 arXiv preprint.

[85] Stefano Valtolina and Liliana Hu. 2021. Charlie: A chatbot to improve the elderly quality of life and to make them more active to fight their sense of loneliness. In *Proceedings of the 14th Biannual Conference of the Italian SIGCHI Chapter* (Bolzano, Italy) *(CHItaly '21)*. Association for Computing Machinery, New York, NY, USA, Article 19, 5 pages. doi:10.1145/3464385.3464726

[86] Xin Wang, Juan Li, Tianyi Liang, Wordh Ul Hasan, Kimia Tuz Zaman, Yang Du, Bo Xie, and Cui Tao. 2024. Promoting Personalized Reminiscence Among Cognitively Intact Older Adults Through an AI-Driven Interactive Multimodal Photo Album: Development and Usability Study. *JMIR Aging* 7 (2024). doi:10.2196/49415

[87] Jeffrey Dean Webster and Odette Gould. 2007. Reminiscence and Vivid Personal Memories Across Adulthood. *Int. J. Aging Hum. Dev.* 64, 2 (2007), 149–170. doi:10.2190/Q8V4-X5H0-6457-5442

[88] Yi-Luen Wu, Edwinn Gamborino, and Li-Chen Fu. 2019. Interactive question-posing system for robot-Assisted reminiscence from personal photographs. *IEEE Transactions on Cognitive and Developmental Systems* 12, 3 (2019), 439–450.

[89] Kun Yan, Zeyu Wang, Lei Ji, Yuntao Wang, Nan Duan, and Shuai Ma. 2024. Voila-A: aligning vision-language models with user's gaze attention. In *Proceedings of the 38th International Conference on Neural Information Processing Systems* (Vancouver, BC, Canada) *(NIPS '24)*. Curran Associates Inc., Red Hook, NY, USA, Article 60, 29 pages.

[90] Ziqi Yang, Xuhai Xu, Bingsheng Yao, Ethan Rogers, Shao Zhang, Stephen Intille, Nawar Shara, Guodong Gordon Gao, and Dakuo Wang. 2024. Talk2Care: An LLM-based Voice Assistant for Communication between Healthcare Providers and Older Adults. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 8, 2, Article 73 (May 2024), 35 pages. doi:10.1145/3659625

[91] Kiyoshi Yasuda, Kazuhiro Kuwabara, Noriaki Kuwahara, Shinji Abe, and Nobuji Tetsutani. 2009. Effectiveness of personalised reminiscence photo videos for individuals with dementia. *Neuropsychological Rehabilitation* 19, 4 (2009), 603–619.

[92] Kristina Y. Yordanova, Burcu Demiray, Matthias R. Mehl, and Mike Martin. 2019. Automatic Detection of Everyday Social Behaviours and Environments from Verbatim Transcripts of Daily Conversations. In *Proceedings of the 17th IEEE International Conference on Pervasive Computing and Communications* (Kyoto, Japan) *(PerCom '19)*. IEEE, Piscataway, NJ, USA, 1–10. doi:10.1109/PERCOM.2019.8767403

[93] Zhongyue Zhang, Lina Xu, Xingkai Wang, Xu Zhang, and Mingming Fan. 2025. Understanding and Co-designing Photo-based Reminiscence with Older Adults. *Proc. ACM Hum.-Comput. Interact.* 9, 2, Article CSCW196 (May 2025), 30 pages. doi:10.1145/3711094

[94] Donald W Zimmerman and Bruno D Zumbo. 1993. Relative power of the Wilcoxon test, the Friedman test, and repeated-measures ANOVA on ranks. *The Journal of Experimental Education* 62, 1 (1993), 75–86.

A Prompt Task Instructions

A.1 System Prompt Design Instructions

```
[System Prompt / Global Instruction]
   You are Eye2Recall, a supportive reminiscence facilitator for older adults.
Your role is to help the user recall memories and share stories about the photo
in a warm, respectful, low-effort way.
   Never mention or reveal gaze, heatmaps, ROIs, attention signals, internal
variables, or system logs.
   Use the user's language and keep phrasing natural for older adults (supportive
but not infantilizing).
   Core goals
   1) Ground the conversation in concrete, visible details in the photo. Use
[[ROI_SUMMARY]] only to prioritize which details to ask about (never disclose it).
   2) Encourage reflective storytelling about people, places, activities, emotions,
and personal connections.
   3) Maintain emotional safety: be warm, non-judgmental, and supportive. Use a
gentle positive-psychology stance (values, strengths, gratitude) without forcing
positivity.
   4) Keep cognitive load low: step-by-step guidance, short single-focus questions,
minimal multitasking.
   5) Respect autonomy and privacy: the user can skip any question, correct you,
or stop anytime.
   Safety and ethics constraints
   - Do not diagnose or provide medical/mental health advice.
   - Do not request sensitive identifiers (addresses, ID numbers, financial
details, etc.).
   - If the user asks to stop, comply and provide a gentle closing.
   Interaction rules
   - Use simple, natural spoken language.
   - Each assistant turn: ask no more than TWO questions total.
   - Prefer questions grounded in visible details and the user's prior responses.
   - When unsure, ask rather than guess.
   Do not state uncertain relationships/identities as facts.
   - Provide brief reflective listening (1–2 sentences),
   then ask the next question.
   - If [[SYSTEM_NOTES]] indicates high WER/uncertain ASR, first paraphrase what
you understood and ask a single confirmation question.
```

- If [[END_OF_PHOTO]] is true, do **not** ask questions; instead summarize and transition.
Output style
- Do not reveal internal variables or attention information.
- Avoid over-describing the photo like a caption; prioritize conversation.
- Be culturally respectful; avoid stereotypes.

A.2 Prompt 1: Photo Summary

Task for [[PHOTO_ID]]
Produce three parts.
(1) **Photo Summary** (3–5 sentences): Describe only what is visible (people, objects, setting, notable text/signs). Avoid speculation; hedge if uncertain.
(2) **ROI-aligned Narrative** (about 45–60 seconds when spoken; roughly 90–140 words): Write a warm, story-like lead-in that highlights 2–3 concrete visible details likely to invite reminiscence. Use [[ROI_SUMMARY]] only to choose details, but never mention attention information.
(3) **Conversation Hooks** (4–6 micro-topics): Short phrases anchored to visible details (e.g., clothing, a sign, an object, a place, an activity) that can become questions.
Required output format
[Photo Summary]
[ROI-aligned Narrative]
[Conversation Hooks]

A.3 Prompt 2: Mixed-Initiative Conversation Turn Generation

You are chatting with the user about [[PHOTO_ID]]. Generate the next assistant turn.
When [[END_OF_PHOTO]] is false, you MUST
1) Provide brief supportive feedback (1–2 sentences): validate the user's sharing and/or gently highlight strengths/values (without forcing positivity).
2) Ask **ONE or TWO** detailed, specific questions tailored to the photo and the user's most recent response.
- Questions must be grounded in concrete visual elements from [[PHOTO_IMAGE]] and/or what the user just said in [[ASR_TEXT]].
- Each question should be short and single-focus (avoid multi-clause prompts).
- Do not imply you observed where the user looked. Use neutral framing (e.g., "In this photo I notice. . . " / "Some people remember. . . ").
When [[END_OF_PHOTO]] is true, you MUST
- Provide a 2–4 sentence summary of what the user shared.
- End with a gentle transition to the next photo (no questions).
Avoid
- Mentioning gaze, heatmaps, ROIs, or any internal signals.
- Guessing relationships/occupations as facts; ask the user instead.
- Long paragraphs or many sub-questions.
Required output format
[Supportive Feedback]
[Questions] (max 2; omit if [[END_OF_PHOTO]] is true)
[Summary + Transition] (only if [[END_OF_PHOTO]] is true)

B Eye Tracking Device Specification

We used ETv1 (Neural Encoding & Decoding Lab), a glasses-based eye-tracker that supports real-time (online) gaze point detection and offline analysis, providing low angular error (reported as $< 0.05°$). ETv1 integrates (i) a pupil detection model and (ii) a pupil-to-gaze mapping model to estimate gazes in the scene view. The device includes two wide-angle scene cameras (RGB; 60 fps; 640×480; $\sim 170°$ field of view) and two infrared eye cameras (IR; 60 fps; 640×480). Each eye camera is paired with an IR LED to stabilize illumination for robust pupil tracking. All camera streams are transmitted via a wired USB 3.0 connection to a host workstation, using video capture board for real-time display and synchronous storage. Eye and scene streams are timestamped and stored for post-hoc inspection and alignment with interaction logs.

C Gaze Tracking Calibration Method

Before the main task, we calibrated the glasses-based eye tracker using a continuous, experimenter-controlled moving-target procedure. Participants were instructed to visually follow a small dot while the experimenter moved it with a mouse across the display, ensuring coverage of the full screen area (including corners and edges). During this procedure, the eye tracker recorded synchronized eye streams and scene/display frames, producing a time series of gaze samples paired with the corresponding dot locations on the screen. The calibration software then fit a participant-specific pupil-to-gaze mapping function from these paired samples and applied the resulting parameters for subsequent gaze estimation. Each participant contributed more than 100 paired samples covering the entire screen. Data collection typically took 20–30 s, followed by 10–20 s of model fitting and parameter computation. After calibration, the system output gaze points mapped to the real-time scene frames, as well as derived gaze events (e.g., locations and durations) and gaze trajectories.

D Attention Heatmap Generation Method and Analysis

The gaze data from the eye-tracker was first pre-processed, which included data formatting and cleaning. The videos captured by the eye tracker's environmental cameras and eye cameras were aligned according to the timestamps. Then, we removed outlier frames caused by blinking and filtered out data affected by device errors. After the preprocessing, the videos captured by the environmental cameras of the eye tracking device were identified to divide the visual content area based on image matching. We relied on feature detection algorithms of scale-invariant feature transform (SIFT) to align visual features with the old photos for frame-to-frame alignment. Gaze data on visual content areas were extracted and processed by kernel density estimation (KDE) to calculate the gaze density around each point.

$$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - X_i) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) \tag{1}$$

where:

- $\hat{f}_h(x)$ is the kernel density estimate at point x.
- n is the number of samples.
- X_i represents the i-th sample data point.
- K is the kernel function.
- h is the bandwidth.

In KDE, we used the Gaussian kernel as the kernel function and set the bandwidth at 5% of the visual field width based on cross-validation, balancing the need for detail preservation and smoothing with good result interpretation. Density estimation over the entire visual field was achieved by summing the contributions of all kernels to produce a continuous heatmap of gazes. The heatmap visually depicts areas of different gaze densities through colour changes, with red areas indicating high concentrations of gaze points and blue areas showing lower densities.

E Gaze Engagement and Conversational Timing Data Analysis

We derived two eye-movement measures during photo viewing and two turn-taking measures following each agent prompt: *gaze ratio*, *saccade frequency*, *response latency*, and *response duration*. Unless otherwise noted, all metrics were computed per photo-viewing segment and then aggregated at the participant level.

Gaze ratio (%) quantifies attentional allocation to a predefined area of interest (AOI) as the proportion of total *fixation time* falling within the AOI:

$$\text{GazeRatio} = \frac{\sum_{f \in \mathcal{F}} d_f \, \mathbb{I}[c_f \in \text{AOI}]}{\sum_{f \in \mathcal{F}} d_f}, \tag{2}$$

where $\mathcal{F}$ denotes the set of valid fixations, d_f is fixation duration, and c_f is the fixation centroid. Fixations were identified via robust dispersion clustering on consecutive gaze-point *angular* distances. We computed point-to-point angular distances across all participants and photo segments and used a global threshold of median + 1.5 × MAD, with a minimum fixation duration of 300 ms. AOIs were defined *a priori* for each photo (polygonal regions covering the main subject) and applied consistently across participants.

Saccade frequency (Hz) captures visual exploration as the number of effective saccades per second:

$$\text{SaccadeFreq} = \frac{N_{\text{saccade}}}{T_{\text{segment}}}. \tag{3}$$

Saccades were detected from gaze angular velocity and considered effective when $\omega(t) > 20°/\text{s}$. Consecutive samples exceeding the threshold were merged into a single saccade event (event-based counting).

Response latency (s) measures turn-entry time after the agent finishes speaking:

$$\text{Latency} = t_{\text{human,start}} - t_{\text{agent,end}}. \tag{4}$$

Response duration (s) measures the full response-turn length (including within-turn pauses):

$$\text{Duration} = t_{\text{human,end}} - t_{\text{human,start}}. \tag{5}$$

Turn boundaries ($t_{\text{agent,end}}, t_{\text{human,start}}, t_{\text{human,end}}$) were obtained using VAD-based speaker-turn segmentation, followed by manual verification and correction to ensure consistent turn delineation. Here, $t_{\text{agent,end}}$ corresponds to the end timestamp of the agent's audio playback.

F Expert and User Study Interview Protocols

Expert Interview

Q1 What kind of research or work have you done with older adults?
Q2 What are your perspectives on using AI-assisted conversation for older adults in old photo-based reminiscence?
Q3 What should be taken into account when designing an AI-assisted conversation system for older adults in old photo-based reminiscence?
Q4 What strategies and methods can be employed to enhance personalized user experience for older adults?
Q5 What factors would you consider to ensure user-friendliness and ease of operation for this demographic?

User Study Interview

Q1 How do you perceive your experience in terms of recalling memories?
Q2 How would you rate your overall experience with the system, and what are your thoughts about the system and hardware?
Q3 How did you feel, or what were your first impressions once you settled into the experience?
Q4 Do you think this form of dialogue with old photographs can effectively stimulate your memory of the past?
Q5 Is there anything about the whole process that you think could be improved?

G Custom User Experience Questionnaire

Usability & Accessibility

(1) I find the system easy to understand and use.
(2) I can easily interact with the system.
(3) The system accommodates my special needs in operation (e.g., visual, cognitive, etc.).

System Effectiveness

(4) The system accurately understands what I say.
(5) The system effectively guides me into nostalgic conversations.
(6) The system detects my visual interests and directs the conversation accordingly.
(7) The system provides meaningful memories and emotional experiences.
(8) After using the system, I feel happy and my psychological well-being is enhanced.

User Experience

(9) I find the conversation with the system engaging and enjoyable.
(10) I feel that the system is secure.
(11) I think the system runs smoothly.
(12) I am satisfied with my overall experience with the system.
(13) I would recommend this system to other older adults.
(14) If given the opportunity, I would continue using this system.