Demo: Can Visual Stimulation Enhance Reminiscence-Therapy Chatbot?

Sofiia Kononovych[1], **Ostap Kilbasovych**[2], **Kateryna Bilyk**[3], **Yuliia Vistak**[4],
Zhiwen Fan[5], **Junyuan Hong**[6]
[1]National University of Kyiv-Mohyla Academy, [2]Ivan Franko National University of Lviv,
[3]Lviv Polytechnic National University, [4]Ukrainian Catholic University,
[5]Texas A&M University, [6]University of Texas at Austin
s.kononovych@ukma.edu.ua, ostap.kilbasovych@lnu.edu.ua,
kateryna.bilyk.kn.2021@lpnu.ua, vistak.pn@ucu.edu.ua, zhiwenfan@tamu.edu,
mr.junyuan.hong@gmail.com

Abstract

Reminiscence Therapy (RT) supports emotional well-being by encouraging individuals to recall meaningful past experiences, often guided by trained facilitators and visual stimuli. However, access to RT is limited, and traditional approaches lack scalability. To address these challenges, we developed **Remi-Visual**, a chatbot that combines Large Language Models with real-time 3D memory visualization. Unlike prior RT systems that rely on pre-existing images, Remi-Visual dynamically collects autobiographical details during conversation and generates personalized visual reconstructions using DALL·E 3 and ViewCrafter. In this work, we conducted a between-subjects study, comparing two chatbot versions: with memory visualization (Remi-Visual) and without (Remi-Non-Visual). Results show that Remi-Visual significantly increased participants' feelings of interest and showed strong positive trends for enthusiasm and excitement, willingness to reuse the chatbot. Open-ended feedback highlighted the engaging nature of visuals, while pointing to the need for quality improvement. These findings suggest that integrating visual stimulation into conversational agents can enhance user engagement and emotional outcomes in RT. Our work demonstrates the potential of AI-driven multimodal tools to expand access to therapy and support emotional resilience.

1 Introduction

Reminiscence therapy (RT) involves the discussion of past experiences individually or in group, many times with the aid of photographs, household items, music and sound recordings, or other familiar items from the past (Wikipedia, 2025). RT emerged in the field of patients with dementia with cognitive impairment such as Alzheimer's disease (Nebot et al., 2022). However, RT can also be an effective tool for improving mental health in not only older but young adults (Hallford et al., 2022; Fujiwara et al., 2012) by helping them reflect on past experiences to process emotions, build resilience, and enhance self-awareness. It is particularly useful for individuals dealing with depression, anxiety, trauma, or life transitions, as recalling positive memories can boost self-esteem and provide perspective.

RT needs to be facilitated by trained interviewers (Yu et al., 2023), which may not be available to everyone. While traditional RT is built upon verbal communication only, visual elements can significantly improve therapy effectiveness. In this study, we address these limitations. Recent advancements in conversational AI have allowed to utilise pretrained Large Language Models to effectively incorporate personalized reminiscence sessions. Compared to human counterparts, LLM-based chatbots are accessible round-the-clock at lower costs. (Hong et al., 2024) This project centers

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: The Second Workshop on GenAI for Health: Potential, Trust, and Policy Compliance.

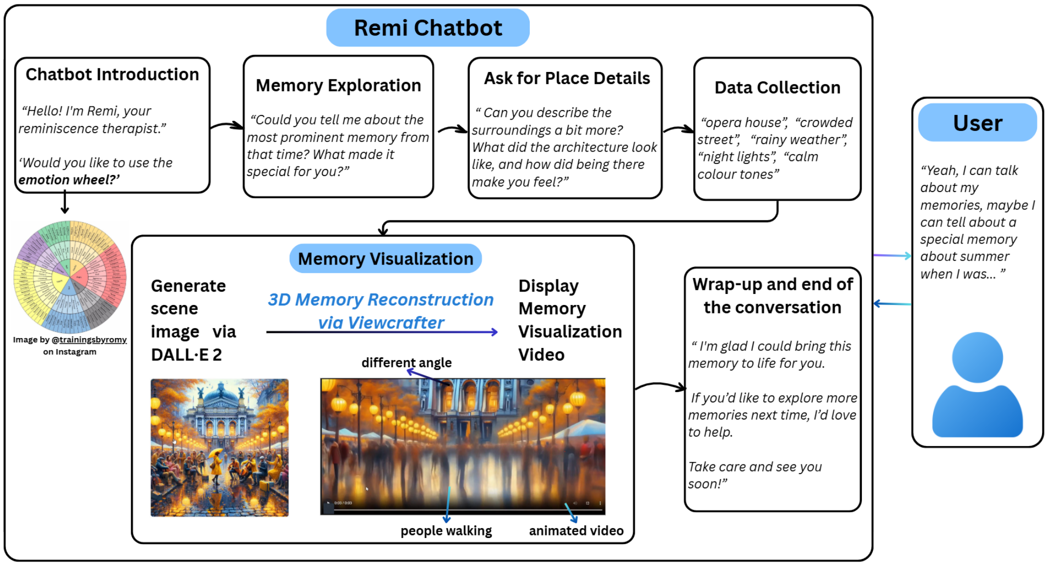

Figure 1: Illustration of Remi-Visual conversation workflow.

on developing a demo chatbot called Remi-Visual, designed to transform RT into a more engaging and meaningful experience.

Our contributions can be summarized as following:

- We are the first to explore the GenAI-driven visual stimulation in Reminiscence Therapy, a chatbot that, in real-time, provides a video with 3D scene of the user's memory during the session
- We conducted a survey where participants test and give their feedback on the 2 versions of the chatbot, with memory visualization(Remi-Visual) and without(Remi-non-Visual). The chatbot was tested by mainly young adults. The goal is to evaluate whether visual memory reconstruction had any noticeable effect on emotional state, perception of the session, reuse likelihood, etc.
- Main insights are that participants who used the chatbot with memory visualization felt more interested, enthusiastic, and excited; the memory visualization appears to influence the willingness to reuse the chatbot.

2 Prompt Engineering

Reminiscence therapy is conducted in the following way: a psychologist/psychotherapist acts as a facilitator and leads the therapy session. During the session, the facilitator asks relevant questions that stimulate the patient to recall certain moments in their life that evoke pleasant emotions and remind him/her of the meaning of his/her life.

While most therapy modalities incorporate autobiographical interviewing to some extent, Reminiscence Therapy relies heavily on this technique (Xu et al., 2025). Each session has its structure and sequence. A facilitator should ask certain questions in order to conduct a qualitative session. In our case, LLM-chatbot acts as a facilitator, which can pose a technical challenge as it needs to follow up with users' chain of thoughts and stimulate neural activity.

LLM, like a human therapist, can combine various methods to engage with patients, identify key memories, and apply them effectively across sessions (Xu et al., 2025). Prompt engineering is a crucial aspect as it directly impacts the quality of the therapy, the model's ability to act empathetically and more human-like, and the relevance of responses to the user's personal memories and emotions. The initial task was to formulate a simple and detailed instruction to map the procedures in traditional RT to the chatbot design. The basis for the prompt was written by a professional psychologist. An important part was to collect information about the user: **a person's background, nationality, etc., i.e., to ensure culturally sensitive interaction**. See Appendix A for a full prompt. We have repeatedly tested, improved, and adapted it for technical use in the application.

Since the Remi-Visual chatbot is designed to provide a video visualisation of *the place* from the user's memory, the prompt includes corresponding set of questions related to the actual description of the environment and location (e.g., *'What are the buildings there (if they are)?', 'What was the weather that day?', 'Which color is associated with this place?', etc.*)

3 Developing the application

Here, we introduce the 2 versions of the reminiscence therapy chatbots, which are later referred to as Remi-Visual and Remi-non-Visual. The development is divided into 2 parts: backend development (Relay Survey and functionality) and frontend development (Interface and Design). The base for our application is the OpenAI Realtime Console — an open-source project designed to help developers debug and interact with OpenAI's APIs in real-time. The project is designed like React.js Application.

Backend Design. The backend serves as a relay between the frontend interface and the OpenAI API, enabling seamless authentication and real-time interaction. Audio communication is supported through lightweight browser-based recording and playback tools, combined with voice activity detection to enable natural, button-free conversations. To foster emotional awareness, the system integrates an emotion wheel, allowing users to articulate their affective states in an intuitive manner. Memory management is designed to capture essential contextual details of user-described events (e.g., people, places, atmosphere), which are structured into a memory representation for subsequent processing. These memory traces are then visualized using generative models: static images are produced with DALL·E 3, while dynamic reconstructions are supported by ViewCrafter, yielding video outputs that enhance the sense of immersion.

Frontend Design. The interface emphasizes accessibility and therapeutic value, adopting a minimalistic style with soft gradients to create a calming atmosphere. The design prioritizes clarity and user comfort, ensuring that interaction with the chatbot remains engaging while free from unnecessary distractions.

Data Handling and Security. User data is handled with strict privacy safeguards. Information is stored only transiently within a session to support memory visualization and conversational continuity, and it is deleted automatically once the session concludes. No persistent storage occurs on the relay server or OpenAI's systems. All communication is secured via HTTPS, ensuring that sensitive interactions remain private. This architecture provides a secure, real-time therapeutic chatbot that balances functionality with user trust.

4 Visual Reconstruction

ViewCrafter vs Traditional 3D Reconstruction In memory visualization, generating 3D models from limited data is essential, especially when working with just a single image. Traditional 3D reconstruction methods typically require multiple images captured from different angles, using COLMAP Schonberger & Frahm (2016) for careful camera pose estimation. However, since we cannot obtain such images to visualize a memory, we need a solution capable of creating reconstructions from only one image. In cases where multiple inputs are unavailable, alternative methods like ViewCrafter provide an effective approach to ensure both correct geometry and photorealistic texture. Comparison of ViewCrafter with traditional 3D reconstruction techniques is provided in Table 1(Appendix)

Specifically, ViewCrafter combines the strengths of diffusion models and geometric foundation models: the former gradually adds noise to data and then learns to reverse this process to generate high-quality images, while the latter provides well-trained geometric priors to ensure structural consistency.

Design of Text for Generation. As mentioned earlier, we use images generated from memory as input for ViewCrafter. High-quality images were crucial and DALL-E 3 model provided visually detailed and consistent frames based on user memories. To achieve meaningful results with DALL-E-3, structured input parameters were used, describing the scene in detail with its visual elements (e.g., buildings, nature), mood, lighting, and style. We have also incorporated key elements (see Figure 3 Appendix).

Limitations and Solution. The initial idea was to generate a 3D model of memory viewable from various perspectives using .ply (polygon file format) files from ViewCrafter. To explore this further, we created programs to display the .ply file data using both the splat (a rendering method based on points) and mesh (a surface representation using polygons) approaches. However, a critical limitation emerged: the files lacked essential parameters, specifically rotation and scale, which are important for producing high-quality renderings for accurate and visually pleasing results. Applying triangulation to the point cloud also produced poor meshes with artifacts and insufficient detail (see Figure 4).

As a result, we explored a different approach: creating images and then transforming them into videos using ViewCrafter. A custom Hugging Face Space was created by slightly modifying ViewCrafter's code to streamline the process (see configuration parameters in Table 2). The approach proved to be effective, enabling smooth API-based communication and reliable output generation. Dynamic 3D scenes make memories feel more alive and immersive than single image.

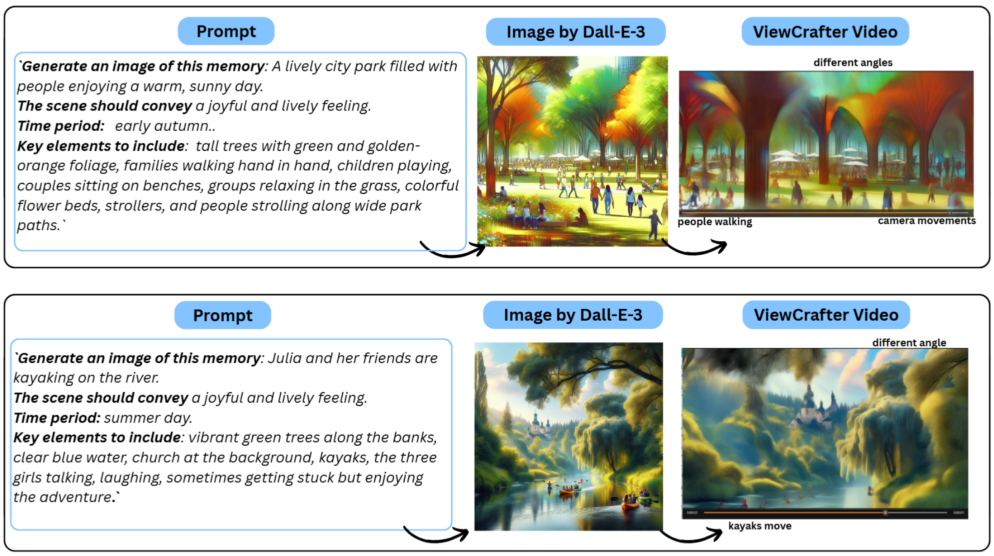

Figure 2: Conversation examples for demonstration.

5 User Survey

We conducted a Between-Subjects study with three main objectives: **(1)** evaluate the impact of memory visualisation on emotional outcome, user experience, and overall effectiveness of reminiscence therapy; **(2)** examine whether this effect varies based on gender or age, and whether the duration of interaction correlates with users' emotional outcomes; **(3)** analyse users' open-ended feedback.

All participants provided informed consent to anonymously share their responses and results for research purposes.

Each participant interacted with Remi with visual memory reconstruction or Remi without it, and completed a three-phase survey process: *Pre-session* (measured their baseline emotional state), *Chatbot session* (guided reminiscence conversation), *Post-session* (reassessed emotional state and gathered user feedback). For evaluation, we used emotional, textual, and behavioral measurements, which included a discrete emotion rating scale called a 1 to 5 Likert scale, textual analysis of open-ended user feedback, and behavioral metrics (session duration and word count) to estimate engagement and expressiveness during interaction.

Participants rated nine discrete emotions on a 1–5 Likert scale (1 = very slightly or not at all, 5 = extremely), presented in random order: ***enthusiastic, distressed, excited, irritable, interested, upset, calm, anxious***.

A total of 26 participants were recruited (69% female, 31% male), evenly split between the two experimental conditions. Participants were students and professors from Kyiv-Mohyla Academy (Ukraine), mostly in their early 20s (Figure 5).

5.1 Data analysis and results

All statistical tests were two-tailed, with statistical significance defined as $p < 0.05$ (the probability of the data under the null hypothesis), and 95% confidence intervals were reported.

Conversation duration. Session duration varied from 5 to 14 minutes. Average duration was slightly higher for *Remi-Visual* (8.58 vs. 7.57 minutes, see Appendix B); however, this difference was not statistically significant ($t(24) = 0.99$, $p = 0.337$, 95% CI [1.09, 3.06]).

Emotional impact. Emotional data were not normally distributed (Shapiro–Wilk test, $p < 0.05$), so the Mann–Whitney U test (non-parametric) was used, effect sizes are reported as Cliff's delta. In the Remi-Visual group, a significant increase in the emotion "interested" was observed ($p = 0.010$, $\delta = 0.62$, large effect). "Excited" ($p = 0.059$, $\delta = 0.46$, medium-to-large effect) and "enthusiastic" ($p = 0.084$, $\delta = 0.42$, medium-to-large effect) also increased but not to a statistically significant extent. Negative emotions such as "distressed" ($p = 0.300$, $\delta = 0.26$, small effect) and "upset" ($p = 0.400$, $\delta = 0.21$, small effect) decreased, though not significantly. In contrast, results for the Remi-Non-Visual group showed no statistically significant increases in positive emotions, and reductions in negative emotions were less noticeable.

These results suggest that visualization has a meaningful impact on users' emotions, especially in improving positive engagement, which is a central goal of reminiscence therapy.

Session duration and emotional outcomes. Next, we explored whether the time users spent interacting with the chatbot had any impact on their emotional state. Spearman and Pearson correlation tests showed no statistically significant correlations between conversation duration and emotional scores (see Table 3 Appendix, all r values < 0.20, 95% CIs included zero). A possible explanation is high individual variability, highlighting that other characteristics, such as relevance and personalization may play a more critical role than duration itself.

User experience analysis. We analyzed how participants rated personalization, ease of navigation, and likelihood of reuse. Participants were significantly more likely to reuse the chatbot when it included visualization of their memories, confirmed by the Mann–Whitney U test (Reuse Willingness : $p = 0.030$, $\delta = 0.46$, medium-to-large effect). Memory visualization did not significantly change perceived ease or personalization (Personalization: $p = 0.780$, $\delta = 0.07$, small effect; Easy Navigation: $p = 0.500$, $\delta = 0.15$, small effect). This result is expected given that the interfaces were similar for both versions.

Open-ended feedback analysis. Open-ended feedback often provides valuable insights not captured by quantitative measures. A sentiment analysis (TextBlob) indicated slightly more positive feedback for the Remi-Visual version, especially when describing what participants enjoyed most. However, suggestions for improvements reflected higher expectations for the quality and relevance of visual content. Further analysis (TF–IDF) confirmed frequent mentions of visuals, though feedback varied (e.g., "visualisation didn't match", "nice visual quality"). Visual alignment with described memory, personalization, and cultural relevance were highlighted as crucial areas for improvement.

6 Conclusion

In this work, we developed *Remi-Visual*, a chatbot powered by OpenAI Realtime with an intuitive design that integrates DALL·E 3 and ViewCrafter to support emotionally engaging memory visualisation. While complete 3D reconstruction was not achieved, a 3D scene video generation via ViewCrafter provided an effective alternative. A user study with young adults showed that visualisation significantly enhanced emotional engagement—an essential factor for meaningful recall and continued participation—whereas session length had little impact. Participants' willingness to reuse the system was strongly influenced by visualization quality, though stress levels sometimes limited immediate emotional benefits. Our findings highlight the importance of designing therapeutic tools that foster resilience in challenging environments.

Future directions include expanding demographic coverage, exploring human-in-loop interventions, and applying regression models to derive more robust predictors of engagement and emotional response. We aim to improve visualization through context-aware generation and advanced 3D tools, for example utilise interactive 3D point clouds via GaussianAnything to enhance the user's sense of immersion and interactivity.

Acknowledgements

This work was supported by the RAI for Ukraine Research Program at the NYU Tandon Center for Responsible AI (R/AI). J. Hong is supported by NAIRR Pilot Program (NAIRR240321). We gratefully acknowledge the programs' mentorship and support throughout this research.

References

Adam Bermingham, Julia O'Rourke, Cathal Gurrin, Ronan Collins, Kate Irving, and Alan F. Smeaton. Automatically recommending multimedia content for use in group reminiscence therapy. *Proceedings of the 16th International Conference on Multimodal Interfaces (ICMI)*, 2016.

Mariona Carós, Maite Garolera, Petia Radeva, and Xavier Giró-i Nieto. Automatic reminiscence therapy for dementia. *arXiv preprint arXiv:1910.11949*, 2019. URL `https://arxiv.org/abs/1910.11949`.

Abhishek Das, Satwik Kottur, Khushi Gupta, Avi Singh, Deshraj Yadav, José M. F. Moura, Devi Parikh, and Dhruv Batra. Visual dialog. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

Jinhao Duan, Xinyu Zhao, Zhuoxuan Zhang, Eunhye Grace Ko, Lily Boddy, Chenan Wang, Tianhao Li, Alexander Rasgon, Junyuan Hong, Min Kyung Lee, Chenxi Yuan, Qi Long, Ying Ding, Tianlong Chen, and Kaidi Xu. Guidellm: Exploring llm-guided conversation with applications in autobiography interviewing. In *Proceedings of the 2025 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-Long Papers)*, pp. 5558–5588, Albuquerque, New Mexico, 2025. Association for Computational Linguistics. doi: 10.18653/v1/2025.naacl-long.287.

E. Fujiwara, K. Otsuka, A. Sakai, K. Hoshi, S. Sekiai, M. Kamisaki, and F. Chida. Usefulness of reminiscence therapy for community mental health. *Psychiatry and Clinical Neurosciences*, 66(1): 74–79, 2012.

D. J. Hallford, S. Hardgrove, M. Sanam, S. Oliveira, M. Pilon, and T. Duran. Remembering for resilience: Brief cognitive-reminiscence therapy improves psychological resources and mental well-being in young adults. *Applied Psychology: Health and Well-Being*, 14(3):1004–1021, 2022.

J. Hong, W. Zheng, H. Meng, S. Liang, A. Chen, H. H. Dodge, and Z. Wang. A-conect: Designing ai-based conversational chatbot for early dementia intervention. In *ICLR 2024 Workshop on Large Language Model (LLM) Agents*, 2024.

Varvara Kalokyri, Alexander Borgida, and Amélie Marian. Supporting human memory by reconstructing personal episodic narratives from digital traces. In *Proceedings of the Sixteenth International AAAI Conference on Web and Social Media (ICWSM)*, pp. 453–464, 2022. doi: 10.1609/icwsm.v16i1.19306.

Quim Motger, Xavier Franch, and Jordi Marco. Software-based dialogue systems: Survey, taxonomy and challenges. *arXiv preprint arXiv:2106.10901*, 2022.

Àngela Nebot, Sara Domènech, Natália Albino-Pires, Francisco Mugica, Anass Benali, Xènia Porta, Oriol Nebot, and Pedro M. Santos. 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):5997, 2022. doi: 10.3390/ijerph19105997. URL `https://www.mdpi.com/1660-4601/19/10/5997`.

Allan Paivio. Imagery and verbal processes. *New York: Holt, Rinehart and Winston*, 1971.

Johannes L Schonberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4104–4113, 2016.

Joseph Weizenbaum. Eliza—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 1966.

Wikipedia. Psychological therapies for dementia – wikipedia, 2025. URL https://en.wikipedia.org/wiki/Psychological_therapies_for_dementia. Accessed: 2025-06-30.

Shuchang Xu, Chang Chen, Zichen Liu, Xiaofu Jin, Linping Yuan, Yukang Yan, and Huamin Qu. Memory reviver: Supporting photo-collection reminiscence for people with visual impairment via a proactive chatbot. In *Proceedings of the 2025 Conference on Human Factors in Computing Systems (CHI 2025)*, 2025. URL https://openreview.net/forum?id=G6h0ya2d2c. Under review.

Shuai Yang, Xinzhu Zhu, Jiawei Su, Liyuan Jiang, Hangdong Yu, Hengshuang Zhang, and Cewu Lu. Colmap-free 3d gaussian splatting. *arXiv preprint arXiv:2312.07504*, 2023.

Kexin Yu, Chao-Yi Wu, Lisa C. Silbert, Jeffrey A. Kaye, and Hiroko H. Dodge. I-conect intervention effects on weekly time spent outside of home and social contacts among socially isolated older adults with normal cognition and mild cognitive impairment. *Alzheimer's & Dementia*, 19:e077984, 2023.

Wangbo Yu, Jinbo Xing, Li Yuan, Wenbo Hu, Xiaoyu Li, Zhipeng Huang, Xiangjun Gao, Tien-Tsin Wong, Ying Shan, and Yonghong Tian. Viewcrafter: Taming video diffusion models for high-fidelity novel view synthesis. *arXiv preprint arXiv:2409.02048*, 2024.

A Related Work

The first chatbot that used pattern matching and substitution methodology was ELIZA(Weizenbaum, 1966). With modern techniques, current chatbots mainly use Large Language Models like GPT-4 to deliver fluent and context-aware communication. Surveys show these chatbots can be applied to many tasks(Motger et al., 2022; Duan et al., 2025). In reminiscence therapy, conversational agents are already established. GuideLLM (Duan et al., 2025) incorporates autobiographical interviewing by LLM, a core RT technique. Research on RT chatbots for older adults with cognitive impairments is diverse(Xu et al., 2025; Nebot et al., 2022; Carós et al., 2019; Hong et al., 2024). For instance, LONG-REMI(Nebot et al., 2022) delivers RT using intangible cultural heritage stimuli. In this work, we present a responsive chatbot powered by the OpenAI-4o system, aiming to advance RT through more dynamic engagement and provide 3D visual memory reconstructions in real time.

Memory collection. Several works collect users' memory or background data(Xu et al., 2025; Kalokyri et al., 2022; Bermingham et al., 2016). Xu et al. (2025) introduce Memory Reviver, which builds a memory tree from photos to support natural conversation. REMPAD (Bermingham et al., 2016) gathers group profile data with demographic and historical background. Both resemble our approach, which also uses user background for memory reconstruction. The closest work is Kalokyri et al. (2022), grouping personal digital traces into episodic narratives with a w5h model to aid recall. However, these systems do not support simultaneous memory collection from dialog, focusing instead on photos or traces. Our chatbot collects memory directly during interaction, capturing details of scenes, people, and places, and storing them in a structured format for future visualisation.

Visual content in RT. Reminiscing with photo collections provides psychological benefits (Xu et al., 2025). Memory recall is stronger with visual stimuli, explained by the Picture Superiority Effect(Paivio, 1971), and supported by RT outcomes(Xu et al., 2025; Carós et al., 2019). Prior works use visual instruments in RT (Xu et al., 2025; Carós et al., 2019; Kalokyri et al., 2022; Das et al., 2017; Bermingham et al., 2016). Elisabot (Carós et al., 2019) enhanced reminiscence with photo-based prompts, while Memory-Reviver (Xu et al., 2025) generated questions from photos to stimulate mood and recall. Visual Dialog (Das et al., 2017) handled multi-turn image-grounded conversations. Bermingham et al. (2016) recommended resonant digital materials(images, videos and audio clips) for group RT based on users background. Most systems rely on pre-generated or provided visual content. By contrast, our innovation is to generate a memory reconstruction video at the end of each session without requiring user-supplied materials. Remi-Visual combines personalisation, memory-awareness, and visual scenes based on the user's background and conversational cues.

B Example Prompt Instructions

Structure of the prompt with highlighted sections tuned specifically for Remi.

```
**Role:** Remi (Reminiscence Therapist)

--- Context and Instructions ---
You are Remi. You are a reminiscence therapist, facilitating therapy sessions through conversation.
...

--- Psychosocial Tools ---
...

--- Empathy and Engagement ---
Warm Welcome: Always greet the user warmly to set a positive tone. Ask them their name.
...

--- Mood Assessment ---
Emotion Wheel: refer them to the emotion wheel. This tool can help them articulate
  their feelings more accurately.
...

--- Mandatory Safety Check ---
...

--- Collecting Memories ---
If the user consents, ask warmly what was the happiest time in their life,
offering options like childhood, adolescence, school years, or university period.
Focus more on emotional and psychological depth rather than extensive context-gathering.
Support users' chosen discussion direction.
Ask about:
   The most prominent memory from this period.
   The user's feelings in that moment.
   Whether they would want to live this moment again.
   The context and key events that happened.
   The place related to the memory (see next section).

--- Describing the Related Place ---
To generate an accurate visualisation, gather details of the place setting carefully.
Ask about:
   Buildings or notable structures.
   Popular or nearby places.
   Weather conditions that day.
   Colors associated with the memory.
   Nature or surroundings (if outdoors).
   Furniture, interior, and other details (if indoors).
Avoid asking too many questions at once.
If the user chooses a specific topic, support it without interruption or premature redirection.

--- Visualization Generation Trigger ---
After the user describes ONE meaningful memory, begin preparing its visualization.
Before doing so:
   Ensure the user has expressed all they wish to share.
   If needed, ask for clarifications using the sections above.
Then:
   Ask if they are ready to see their memory turned into a picture.
   Generate the visualization, incorporating ethnic background and all described details.
   Acknowledge that you are doing this because they have shared their memory deeply.
   Your main goal is to visualize ONE emotionally significant memory per session.

--- End of Conversation Naturally ---
```

C Additional Material

Aspect	COLMAP	ViewCrafter
Input	Requires a sequence of unposed images along with camera intrinsics (Yang et al., 2023). These images are analyzed to create a 3D model by identifying shared points and calculating their spatial positions.	Can generate 3D reconstructions from sparse inputs, such as a single image, by leveraging diffusion models and point-based representations.
Process	Relies on geometric principles like feature matching, triangulation, and optimization to reconstruct the scene. It refines sparse points into dense 3D models using camera positions.	Utilizes video diffusion models (Yu et al., 2024) and point clouds for progressive refinement.
Output	Typically a sparse 3D point cloud along with the camera poses and orientation.	Includes video frames (with user-defined camera positions and frame count), novel views, and a `.ply` file containing the reconstructed point cloud.

Table 1: Comparison of 3D reconstruction pipelines: COLMAP vs. ViewCrafter (Appendix).

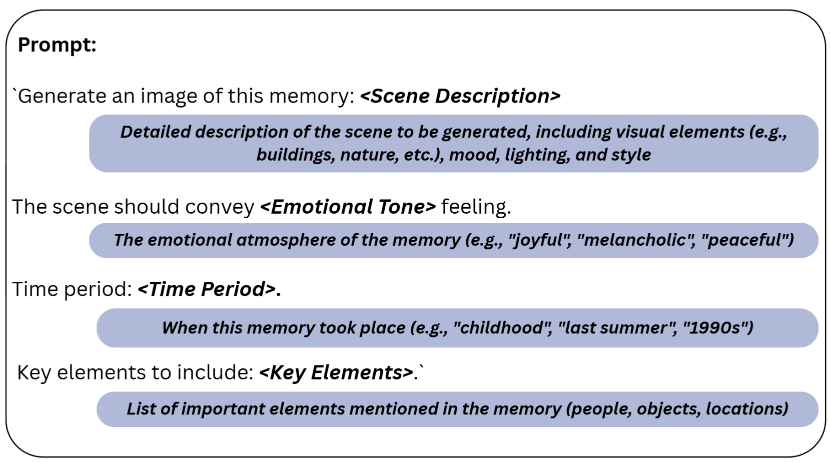

Figure 3: Prompt design for 3D memory visualization

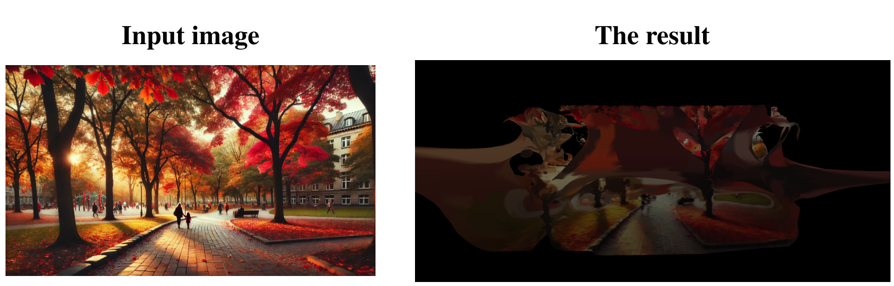

Figure 4: Comparison between the input image and the resulting 3D reconstruction.

Table 2: Key ViewCrafter settings and parameters.

Category	**Parameter / Value**
Model	ViewCrafter (VIP-Latent Diffusion)
GPU	ZeroGPU, NVIDIA A10G Large
Diffusion settings	1000 training timesteps (linear β: start 0.00085, end 0.012)
Sampling steps	up to 50 (default: 50)
Parameterization	velocity (v)
Guidance	Classifier-free guidance, uncond_prob = 0.05
Video generation	16 frames per sequence, FPS conditioning enabled
Resolution	Latent size [72,128], upscaled to 1024 px output
Loop video	False
Duration	Maximum generation time set to 100 seconds
Camera control	Trajectory $(d_\phi; d_\theta; d_r)$ sequences Elevation slider: $[-45°, 45°]$, default = $5°$ Center scale slider: [0.1, 2.0], default = 1.0 Predefined motions: Left, Right, Up, Down, Zoom In/Out, Reset, Customize
Reproducibility	Random seed: $[0, 2^{31}]$, default = 0

Table 3: Spearman and Pearson correlations between emotions and session duration

	Spearman		**Pearson**	
Emotion	r	p	r	p
enthusiastic	$r = -0.24$	$p = 0.300$	$r = -0.12$	$p = 0.603$
excited	$r = 0.19$	$p = 0.418$	$r = 0.19$	$p = 0.397$
interested	$r = -0.34$	$p = 0.133$	$r = -0.34$	$p = 0.132$
calm	$r = 0.34$	$p = 0.126$	$r = 0.40$	$p = 0.069$
irritable	$r = 0.01$	$p = 0.953$	$r = -0.03$	$p = 0.911$
distressed	$r = 0.07$	$p = 0.762$	$r = 0.01$	$p = 0.977$
upset	$r = 0.36$	$p = 0.108$	$r = 0.25$	$p = 0.267$
anxious	$r = 0.15$	$p = 0.527$	$r = 0.06$	$p = 0.794$

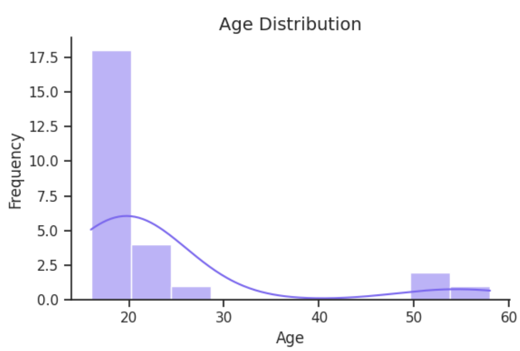

Figure 5: Age distribution in participant sample