A Personalized Embodied Conversational Agent-Based System for Dementia Reminiscence Therapy

Santosh Patapati
santosh@cyrionlabs.org
Dept. of HCI, Cyrion Labs
Dallas, Texas, USA

Trisanth Srinivasan
trisanth@cyrionlabs.org
Dept. of HCI, Cyrion Labs
Dallas, Texas, USA

Rushil Kukreja
rkukreja@cyrionlabs.org
Dept. of HCI, Cyrion Labs
Alexandria, Virginia, USA

Keywords

Embodied Conversational Agent, Dementia, GraphRAG

ACM Reference Format:

Santosh Patapati, Trisanth Srinivasan, and Rushil Kukreja. 2025. A Personalized Embodied Conversational Agent-Based System for Dementia Reminiscence Therapy . In *Proceedings of the 33rd ACM International Conference on Multimedia (MM '25), October 27–31, 2025, Dublin, Ireland.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3746027.3754490

1 Abstract

We introduce an Embodied Conversational Agent (ECA)-based system that delivers daily, personalized reminiscence therapy for people living with dementia, whether at home, in assisted-living, or in nursing-home settings. The system (1) automatically transforms caregiver archives (diary excerpts, letters, posts, or custom memories) into a knowledge graph for memories; (2) retrieves those memories at run-time via Graph-based Retrieval Augmented Generation (GraphRAG) to generate relevant prompts for the ECA; (3) conducts natural, multimodal dialogue powered by a fine-tuned LLaMA-3 13B model; and (4) monitors well-being with a quad-modal cross-attention-based model that tracks 25 Diagnostic and Statistical Manual of Mental Disorders-5 (DSM-5) dimensions of mental health using audio, video, text transcript data, and questionnaire responses. Caregivers can upload new content through a secure interface and preview extracted people/places/dates. They can choose whether the user receives a short daily "Memory Stroll" session (≈5 minutes), a twice-weekly "Deep Dive" session (≈20 minutes) for richer storytelling and cognitive puzzles, or an ad-hoc "Caregiver-Assisted" session that lets caregivers review well-being trends. This integrated system reduces caregiver burden while fostering engagement and emotional well-being in people with dementia.

2 Introduction

Dementia affects over 55 million people worldwide, placing heavy emotional and economic burdens on caregivers [11, 12]. Reminiscence therapy, the guided recall of autobiographical events, has demonstrated benefits for mood, orientation, and social engagement in mild-to-moderate dementia [17, 19]. However, existing digital tools rely on manual slide-show curation, lack interactive dialogue, and provide no longitudinal feedback [2, 5].

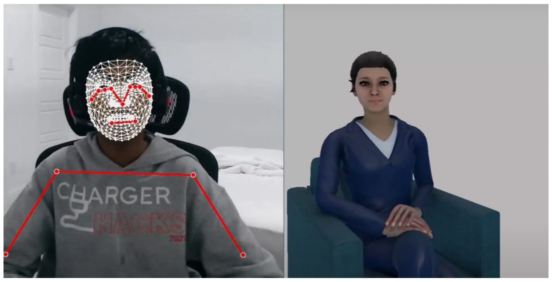

Figure 1: Sample frame from a demo ECA interaction. The ECA maintains a neutral listening position as the user speaks. Visual features are tracked in real-time.

We present the Embodied Reminiscence Agent (ERA), an end-to-end framework that reduces caregiver burden and delivers engaging daily conversations. Our contributions are: (1) Automatic M-KG: diaries, letters, and social-media exports are segmented into "memory atoms", annotated with entities and relations, and stored in a Neo4j knowledge graph; (2) GraphRAG-Based Dialogue: a dual-encoder retriever fetches relevant atoms, which are injected into LLM prompts during ECA interactions; (3) Real-Time Embodiment: speech, affect detection, gesture synthesis, and Unity rendering create an empathetic avatar; and (4) Quad-Modal Mental Health Monitor: audio, vision, text, and questionnaires feed an attention-based network that tracks 25 DSM-5 dimensions and raises caregiver alerts.

3 Methodology

ERA's architecture has six interacting stages. Caregivers upload diary excerpts, letters, photos, or free-text memories. These inputs are segmented into "memory atoms," and a DistilRoBERTa-based Named Entity Recognition (NER) model tags people, places, and dates; a distilled MiniLM-based relation extractor then puts atoms in a Neo4j memory-knowledge graph. During any scheduled session (e.g., daily Memory Stroll or twice-weekly Deep Dive), GraphRAG retrieves top-k atoms based on dialogue state and affect. Those atoms are used for a fine-tuned LLaMA-3 13B response, which the ECA renders in real-time. Mental health questionnaires are administered outside of meetings and feed a quad-modal cross-attention-based model alongside transcripts, audio, and video recordings from meetings to generate well-being scores and alert caregivers.

MM '25, Dublin, Ireland
© 2025 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2035-2/2025/10
https://doi.org/10.1145/3746027.3754490

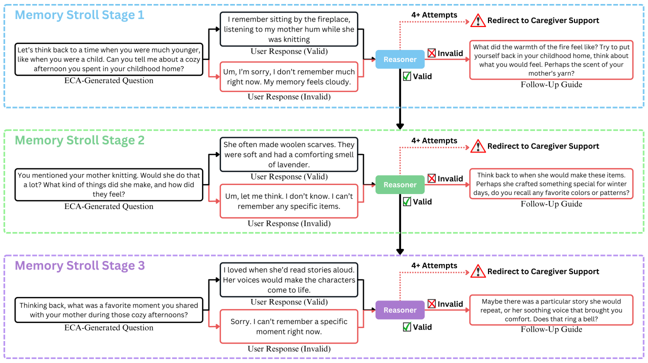
Memory Stroll Stage 1
Let's think back to a time when you were much younger, like when you were a child. Can you tell me about a cozy afternoon you spent in your childhood home?
ECA-Generated Question
I remember sitting by the fireplace, listening to my mother hum while she was knitting
User Response (Valid)
Um, I'm sorry, I don't remember much right now. My memory feels cloudy.
User Response (Invalid)
Reasoner
4+ Attempts
Redirect to Caregiver Support
☒ Invalid
☑ Valid
What did the warmth of the fire feel like? Try to put yourself back in your childhood home, think about what you would feel. Perhaps the scent of your mother's yarn?
Follow-Up Guide
Memory Stroll Stage 2
You mentioned your mother knitting. Would she do that a lot? What kind of things did she make, and how did they feel?
ECA-Generated Question
She often made woolen scarves. They were soft and had a comforting smell of lavender.
User Response (Valid)
Um, let me think. I don't know. I can't remember any specific items.
User Response (Invalid)
Reasoner
4+ Attempts
Redirect to Caregiver Support
☒ Invalid
☑ Valid
Think back to when she would make these items. Perhaps she crafted something special for winter days, do you recall any favorite colors or patterns?
Follow-Up Guide
Memory Stroll Stage 3
Thinking back, what was a favorite moment you shared with your mother during those cozy afternoons?
ECA-Generated Question
I loved when she'd read stories aloud. Her voices would make the characters come to life.
User Response (Valid)
Sorry. I can't remember a specific moment right now.
User Response (Invalid)
Reasoner
4+ Attempts
Redirect to Caregiver Support
☒ Invalid
☑ Valid
Maybe there was a particular story she would repeat, or her soothing voice that brought you comfort. Does that ring a bell?
Follow-Up Guide

Figure 2: State-machine for a Memory Stroll session. Each row shows one conversational phase: orientation (Stage 1), deep recall/enrichment (Stage 2), and positive closure (Stage 3). Black arrows denote the Valid Response Pathway, and red arrows denote Invalid Response Pathways.

3.1 Memory-Knowledge Graph Construction

3.1.1 Segmentation Algorithm. We split uploaded texts into 1-3 sentence "memory atoms" by prioritizing punctuation boundaries ([.!?;:]) and a Gaussian length window (center=50 tokens, $\sigma = 10$). Spans under 20 tokens merge with adjacent spans for adequate context.

3.1.2 Entity & Temporal Extraction. We fine-tune distilroberta-base (82M parameters) [10, 15] via distillation from o3-mini. Training and validation data are drawn from unseen sentences in a 90/5/5 split of English Wikipedia and BookCorpus [3, 21] for joint Person / Place / Date / Event tagging.

3.1.3 Rule-Based Relation Extraction. RegEx templates (e.g., "went to X with Y", "celebrated at X") generate candidate span pairs. We classify each pair with a distilled MiniLM-L6-v2 [18] encoder + MLP, trained via o3-mini distillation on sentences which triggered the RegEx template from the Wikipedia and BookCorpus corpora.

3.1.4 Graph Storage & Retrieval. Memory atoms and entity nodes are stored as nodes in Neo4j. Every text node carries a 768-dimension Sentence-BERT embedding [14] indexed by FAISS [7] for fast approximate nearest neighbor search during GraphRAG retrieval [9].

3.2 Conversational Systems

3.2.1 Session Types. ERA provides three session formats: (1) Memory Stroll (daily, ≈ 5 minutes) orients the user to date/time, elicits short-term recall of positive memories for mood support, and enables quick engagement; (2) Deep Dive (twice weekly, ≈ 20 minutes) guides rich storytelling around an important event and uses simple "who/what/when/where" follow-ups (e.g., "Who joined you?") for cognitive exercise while capturing new details; and (3) Caregiver-Assisted (ad hoc) sessions coach caregivers on how to interact with the patient and review well-being trends. Each session type has its own conversational flow, the conversational flow for Memory Stroll is illustrated in Figure 2.

3.2.2 Cue-Composer & Prompt Templates. : Retrieved memory atoms are formatted (bolding names/dates, adding bridging phrases) and inserted into a fixed LLM prompt template (system instructions + memory block + history + style tokens) for empathetic responses.

3.2.3 General-Purpose Utterance-by-Utterance Interaction Pipeline. The general-purpose utterance-by-utterance ECA interaction pipeline used for every session is executed each time the user presses the speaking button. An "Utterance Segment" is recorded. Audio and video signals are analyzed by lightweight analyzers to detect distress, while a modified Whisper [13] transcribes speech for text validation. Combined multimodal and text analysis generates an ECA response using an LLM (MentalLLaMA-7B-chat [20]), which directs synchronized hand gestures and voice modulation. Real-time audio segments trigger appropriate backchannel behaviors [1]. The resulting lip-sync animations, audio, and nonverbal cues are rendered in a Unity 3D environment. To our knowledge, this is the first ECA framework for psychotherapy capable of delivering real-time empathetic speech and well-timed gestures without human control.

3.3 Quad-Modal Mental Health Monitor

Audio, vision, text, and questionnaires (PHQ-9, GAD-7, and PDS [4, 8, 16]) feed separate encoders (trained on DAIC-WoZ [6]) whose fused cross-attention layers produce 25 DSM-5 scores. Scores above a moderate threshold trigger caregiver alerts on the portal.

3.4 Caregiver Interaction & Support

Caregivers will be able to drag and drop diary excerpts, letters, photos, or type free-text "custom memories" in the secure portal. An instant preview of people, places, and dates is shown before committing to the graph. The same portal visualizes DSM-5 trends and notifies the caregiver regarding any patient mental health concerns.

3.5 Conclusion and Demonstration

We introduced an embodied, retrieval-augmented agent that delivers reminiscence therapy with minimal caregiver burden and real-time empathy cues. For the demo, we present a simplified version of ERA (Figure 1). Attendees upload a pre-written diary entry option to our on-device demo, which immediately segments memories, extracts entities, and updates the knowledge graph. Participants experience a condensed Memory Stroll or Deep Dive session, playing the role of the diarist. Visualizations will display the ECA's real-time thought process, tracked multimodal signals, and more.

References

[1] Pieter A. Blomsma, Gabriel Skantze, and Marc Swerts. 2022. Backchannel Behavior Influences the Perceived Personality of Human and Artificial Communication Partners. *Frontiers in Artificial Intelligence* 5 (2022), 835298. doi:10.3389/frai.2022.835298

[2] Alex Corbett, Joy Myers, and Gregory Roper. 2016. User Perspectives on Digital Reminiscence Tools for People with Dementia. *Journal of Alzheimer's Disease* 54, 3 (2016), 803–811. doi:10.3233/JAD-160121

[3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of NAACL-HLT*. 4171–4186. doi:10.18653/v1/N19-1423

[4] Edna B. Foa, Leslie Cashman, Lisa Jaycox, and Kirsten Perry. 1997. The Validation of a Self-Report Measure of Posttraumatic Stress Disorder: The Posttraumatic Diagnostic Scale. *Psychological Assessment* 9, 4 (1997), 445–451. doi:10.1037/1040-3590.9.4.445

[5] Denis Forbes, Britu Fermin, Coralie Vincent, Alan Leung, Lise Lafortune, and Daniel Lloyd. 2013. Digital Approaches to Supporting Reminiscence Therapy in Dementia Care: A Systematic Review. In *Proceedings of the 2013 ACM International Conference on Health Informatics*. 245–252. doi:10.1145/3624407.3624502

[6] Jonathan Gratch, Andreas Hartholt, George M. Lucas, Georgia Stratou, Stefan Scherer, Amir Nazarian, Douglas Boberg, Albert Rizzo, and Louis-Philippe Morency. 2014. The Distress Analysis Interview Corpus of Human and Computer Interviews. In *Proceedings of LREC*. 3123–3128.

[7] Jeff Johnson, Matthijs Douze, and Hervé Jégou. 2021. Billion-Scale Similarity Search with GPUs. *IEEE Transactions on Big Data* 7, 3 (2021), 535–547. doi:10.1109/TBDATA.2019.2945339

[8] Kurt Kroenke, Robert L. Spitzer, and Janet B.W. Williams. 2001. The PHQ-9: Validity of a Brief Depression Severity Measure. *Journal of General Internal Medicine* 16, 9 (2001), 606–613. doi:10.1046/j.1525-1497.2001.016009606.x

[9] Patrick Lewis, Yuxiang Perez, Anna Piktus, Fabian Petroni, Vladimir Karpukhin, Naman Goyal, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. 2020. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. In *Proceedings of NeurIPS*. 9459–9474.

[10] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach. *arXiv preprint arXiv:1907.11692* (2019).

[11] World Health Organization. 2021. *Dementia: A Public Health Priority.* Technical Report. World Health Organization. https://www.who.int/publications/i/item/9789241565060

[12] Martin Prince, Adelina Comas-Herrera, Martin Knapp, Maëlenn Guerchet, and Maria Karagiannidou. 2015. *World Alzheimer's Report 2015: The Global Impact of Dementia.* Technical Report. Alzheimer's Disease International. https://www.alz.co.uk/research/WorldAlzheimerReport2015.pdf

[13] Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. 2022. Robust Speech Recognition via Large-Scale Weak Supervision. arXiv:2212.04356 [eess.AS] https://arxiv.org/abs/2212.04356

[14] Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence Embeddings Using Siamese BERT-Networks. In *Proceedings of EMNLP-IJCNLP*. 3982–3992.

[15] Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2019. DistilBERT, a Distilled Version of BERT: Smaller, Faster, Cheaper and Lighter. In *Proceedings of the 5th Workshop on Energy Efficient Machine Learning for Embedded Systems*. 1–6. doi:10.18653/v1/W19-4301

[16] Robert L. Spitzer, Kurt Kroenke, Janet B.W. Williams, and Bernd Löwe. 2006. A Brief Measure for Assessing Generalized Anxiety Disorder: The GAD-7. *Archives of Internal Medicine* 166, 10 (2006), 1092–1097. doi:10.1001/archinte.166.10.1092

[17] P. Subramaniam and B. Woods. 2012. The Impact of Reminiscence Therapy on Persons with Dementia: A Systematic Review. *Aging & Mental Health* 16, 8 (2012), 886–898. doi:10.1080/13607863.2012.673152

[18] Xiao Wang, Yukun Wu, Lei Jiang, Jing Xu, and Henry Su. 2020. MiniLM: Deep Self-Attention Distillation for Task-Agnostic Compression of Pretrained Transformers. In *Proceedings of NeurIPS*. 5779–5790.

[19] Barbara Woods, Lisa O'Philbin, Emma Farrell, Abigail Spector, and Martin Orrell. 2018. Reminiscence Therapy for Dementia. *Cochrane Database of Systematic Reviews* 3 (2018), CD001120. doi:10.1002/14651858.CD001120.pub4

[20] Kailai Yang, Tianlin Zhang, Ziyan Kuang, Qianqian Xie, Jimin Huang, and Sophia Ananiadou. 2024. MentaLLaMA: Interpretable Mental Health Analysis on Social Media with Large Language Models. In *Proceedings of the ACM Web Conference 2024 (WWW '24)*. ACM, 4489–4500. doi:10.1145/3589334.3648137

[21] Yukun Zhu, Ryan Kiros, Richard Zemel, Ruslan Salakhutdinov, Raquel Urtasun, Antonio Torralba, and Sanja Fidler. 2015. Aligning Books and Movies: Towards Story-like Visual Explanations by Watching Movies and Reading Books. In *Proceedings of ICCV*. 19–27. doi:10.1109/ICCV.2015.10