Cognitive Computing and Internet of Things, Vol. 165, 2025, 1–10
https://doi.org/10.54941/ahfe1006085

Exploring AI Agents for Reminiscence Therapy in Long-Term Care

Mila Chorbadzhieva, Tim Hulshof, Thijs Tops, Erwin R. Meinders, and Eric Riegen

Mentech Innovation BV, Eindhoven, 5624 CL, The Netherlands

ABSTRACT

With dementia cases rising, long-term care facilities face increasing challenges in providing personalised engagement due to staff shortages. As a result, people with dementia (PwD) and individuals in somatic care are increasingly at risk of social isolation, cognitive under-stimulation, and emotional distress, negatively impacting their overall well-being. Reminiscence therapy (RT) has been shown to improve emotional health and social connection, but its scalability is limited by reliance on caregivers. This study investigates whether Embodied Conversational Agents (ECAs) can facilitate structured RT while reducing caregiver burden. We developed and tested Noa, an AI-driven ECA, in long-term care settings to evaluate engagement, usability, and design factors influencing adoption. The study identified three key design considerations for effective ECA interaction with PwD and people in somatic care: user engagement, avatar design, and interaction modalities. Participants engaged in structured AI-mediated conversations, but speech recognition limitations and AI interruptions disrupted comprehension. Additionally, the system's lack of contextual awareness prevented it from detecting distress, highlighting the need for adaptive dialogue systems to support user well-being. Preferences for avatar design varied significantly, with no universal preference for human-like or abstract avatars, highlighting the need for personalisation. Regarding interaction modalities, wake-word activation failed without assistance, while button-based activation proved more accessible, suggesting the need for multimodal interfaces that accommodate different interaction preferences. Caregivers identified potential for ECAs in companionship and cognitive support, emphasising that integration into existing daily routines is critical for adoption.

Keywords: Embodied conversational agents (ECAS), Reminiscence therapy, Long-term care, Dementia care, Assistive technology, Voice-based interfaces

INTRODUCTION

With more than 280,000 people in the Netherlands (or approximately 1.5% of the population) currently living with dementia—a number projected to rise to 420,000 by 2030 (Ministry of Health, Welfare and Sport, 2020)—long-term care facilities are struggling to provide continuous, personalised care amid growing staff shortages. As resources become increasingly strained, *persons with dementia* (PwD) and those with chronic physical conditions are at greater risk of social isolation, cognitive under-stimulation, and reduced well-being, as there are fewer opportunities for meaningful, consistent interaction.

Reminiscence therapy (RT) is a well-established intervention in dementia care, helping individuals recall and share personal memories to improve emotional well-being, identity, and social connection (Woods et al., 2018). Studies show that RT can reduce agitation, improve social interaction, and enhance quality of life for PwD. However, because RT typically relies on caregiver or family facilitation, its scalability is limited in environments with staffing constraints.

Recent research has explored *Embodied Conversational Agents* (ECAs)—AI-driven virtual characters that simulate human-like conversation and can dynamically present content, such as photos or music—as tools for supporting structured RT. Studies suggest that ECAs could improve access to and engagement with RT by facilitating naturalistic, personalised conversations (Morales-de-Jesús et al., 2021; Rampioni et al., 2021).

Building on this research, we sought to explore whether ECAs could enable RT while reducing reliance on caregiver availability, providing meaningful interaction for PwD. Our prototype was inspired by the AMPER study, which demonstrated the potential of ECAs in guiding reminiscence through structured content selection (Smith et al., 2023). However, while AMPER relies on caregiver mediation, our approach focuses on enabling independent user engagement with the AI system, reducing the need for external facilitation.

This study follows an applied scientific approach within the early development of a potential tool for long-term care. Rather than a purely theoretical investigation, our research aims to generate practical insights that inform real-world implementation. To evaluate this, we developed and tested an ECA for RT, designed to support autonomous conversational interaction. Through exploratory user testing in long-term care settings, we assessed usability, accessibility, and engagement. Discussions with caregivers and participants also revealed broader user needs beyond RT, identifying additional functionalities ECAs could support in dementia care.

PROBLEM STATEMENT AND RESEARCH QUESTIONS

ECAs are being investigated as tools to support PwD, particularly in facilitating structured, meaningful interactions. However, it is unclear whether PwD can and will engage with such systems independently and how different design choices influence their experience.

To address this, we examine PwD's interaction with an ECA and explore the role of design factors. Specifically, we investigate the following research questions:

1. Will people with dementia engage in meaningful and sustained conversations with an ECA?
2. What design factors influence engagement, usability, and accessibility for PwD using an ECA?

To explore these questions, we formulated the following hypotheses:

Hypothesis A (Engagement with ECA): *Users will comprehend and engage in basic conversational interactions with the AI companion.*

Hypothesis B (Avatar Preference): *Users will engage more with a human-like AI avatar due to its familiarity.*

Hypothesis C (Interaction Modality): *Users will prefer using a physical button for the basic device operations over a touch screen interface*. To test these hypotheses, we developed and evaluated an ECA prototype in long-term care settings, as detailed in the following section.

RESEARCH METHODOLOGY

We used a qualitative research approach to develop and test an AI-driven ECA, investigating how PwD interact with it in a simplified form of RT. The study focused on engagement, usability, and key design factors influencing the effectiveness of ECAs in dementia care.

We tested two versions of the ECA prototype across two sessions in long-term care facilities, refining design aspects to evaluate engagement and usability. The prototypes integrated:

- AI-driven conversation models for audio transcription (speech-to-text), text generation (LLM), and audio generation (text-to-speech);
- A hardware setup with wake-word and button-based activation; and
- Three avatar designs: human-like, cartoon-like, and abstract.

Testing involved nine participants, including elderly individuals with dementia, younger individuals with early-onset dementia, and individuals in somatic care. Each session included direct observation of user engagement and interaction, followed by semi-structured caregiver interviews.

To evaluate our hypotheses, we conducted structured observations and video-recorded the sessions for later analysis:

For **Hypothesis A** (Engagement with ECA): we assessed verbal and nonverbal responses, conversational flow, and emotional reactions.

For **Hypothesis B** (Avatar Preference): we evaluated participant reactions, gaze, and sustained attention to the avatar.

For **Hypothesis C** (Interaction Modality): we observed wake-word vs. button activation, focusing on ease and frequency of use. Informed consent was obtained from participants or their legal representatives. All sessions were conducted with a personal assistant caregiver present to ensure participant comfort.

DESIGN OF NOA PROTOTYPE

Approach to Design of the System as a Whole

The design of Noa followed human-centered design (HCD) principles, aiming for engagement and accessibility for PwD and people in somatic care (Norman, 2013). Given the study's seven-week timeframe, we focused on building a functional prototype quickly to enable early real-world testing rather than a fully optimised system. This approach enabled rapid adjustments based on user interactions.

Simplicity and accessibility were central to the design. Research into cognitive and engagement needs for PwD informed the conversational structure, focusing on clear, structured dialogue to reduce cognitive load while maintaining emotional comfort and engagement.

The development process involved deploying an initial prototype, conducting user tests in care facilities, and refining the system. Adjustments were made to the conversation structure, interface design, and hardware configuration for the second prototype, allowing us to test the hypotheses effectively.

Besides this, we also tested with two participants using a classical "*Wizard of Oz*" setup (Kelley, 1984; Nielsen, 1993), where a human tester played the role of Noa via Microsoft Teams (using a built in animated avatar to disguise the tester). This allowed for more control over the conversation than an AI prototype might give, and allowed for testing with very low latency.

Conversation Design

The conversation design for Noa aimed to facilitate meaningful and engaging interactions while ensuring clarity and emotional comfort for users. Following RT principles, Noa used photographs as tangible prompts to evoke positive memories.

Each interaction started with a friendly greeting and a choice between predefined topics based on personal background information, such as hobbies, occupation, or place of residence. Noa then asked if the user wanted view some photos. When the participant responded in the affirmative (all users did), Noa then displayed a relevant photo, described it, and asked structured questions to encourage storytelling and reminiscence. After a series of follow-up questions, a second photo was shown, repeating the process.

The photos were selected prior to the start of the session, such that they matched the interests of the participant, and fit within the conceptual framework of RT. (E.g., a participant who was a liftetime member of a local carnaval association would be shown a photo of his hometown's carnaval parade from years ago.)

To minimise cognitive load while maintaining emotional engagement, the conversation flow was designed with short, direct sentences. Noa dynamically adjusted responses with follow-up questions and gentle or rephrased prompts to sustain engagement. If a user expressed discomfort, Noa politely concluded the conversation to prevent distress.

The design process of the conversation flow was mapped out in diagrammatic form, taking care that it could be easily translated into a system prompt for the LLM (see Diagram 1).

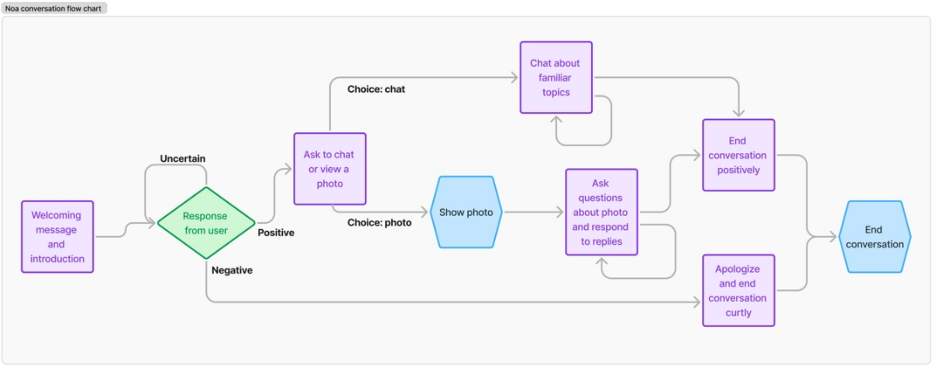

Diagram 1: Simplified schematic of the conversation flow chart for Noa RT interaction. The full schematics considered many more options.

Prompt Design

Noa's system generated structured AI-driven interactions, aiming for clarity and emotional support for users. It used OpenAI's experimental JSON-formatted GPT-4o LLM gpt-4o-2024-08-06 (OpenAI, 2024) as the LLM, enabling Noa to generate structured outputs and execute predefined actions such as displaying images, adapting responses, or closing conversations. Using structured outputs allowed Noa to also operate the app via *special tokens*, for instance, to decide which photo to show, and to end the conversation and return to the start screen.

Because standard GPT-4o responses were too complex for the target group, system prompts were carefully crafted to simplify language while maintaining engagement. Precise wording played a key role in ensuring clear and effective communication. A testing dashboard simulated interactions between Noa and an LLM-generated user persona, allowing rapid iterations to improve phrasing and response clarity before user testing.

The key functional elements of the Noa system prompt were:

1. A brief section with personal information about the participant, including first name, age, gender, former occupation, hometown, and hobbies;
2. A step-by-step written description of the designed conversation flow and decision tree;
3. Instructions for the LLM to use generate special tokens to operate functionalities within the app, such as <show_photo_1>, and <stop_conversation_negative>;
4. Detailed descriptions of all photos;
5. Instructions about behavior, with a strong emphasis on the desired ways of addressing PwD (e.g., to use simple language, ask one question at a time, and not focus too much on things that require short-term memory) and the therapeutic goals of RT.

Figure 1: Stills of the three avatar designs that were used. Left: human-like avatar. Middle: animated otter avatar. Right: abstract pulsating circles avatar.

User Interface and Avatars

Noa's interface was designed for clarity and ease of use, with the avatar as the primary visual element. Subtle facilitator controls enabled conversation flow selection and initiation without distracting users. When displaying a photo, the system minimised the avatar to a corner, naturally shifting the user's focus to the image (see Figure 2). A warm pastel colour scheme was chosen to provide a soothing, non-distracting environment. Animated visual cues indicated listening, processing, or speaking states, enhancing clarity.

Three different avatars were designed for the prototype: one that resembled a real-life human female, one that resembled an animated otter wearing a colorful knitted sweater, and one abstract avatar that consisted of a number of pulsating colored circles (see Figure 1). The animations of all three avatars were simplistic, incorporating only animations for listening, processing, and speaking.

The avatar was intoned using OpenAI's tts-1 model with shimmer and nova voices (for human/abstract and otter avatars, respectively), played at 0.9x normal speech speed.

Hardware

Noa was developed as a tablet-based system, first tested on a laptop and then on a hardware configuration. The main prototype was built around an 8-inch touch screen. To test Hypothesis C we incorporated a modular component with a large mechanical button on a side panel (4cm diameter) for interaction. The hardware casing incorporated wooden elements to create a familiar, inviting appearance, avoiding a clinical look.

The main prototype also incorporated an Intel NUC mini-computer running Ubuntu, housed in a 3D-printed casing. A JBL Bluetooth speaker was added to improve sound quality, and a mobile hotspot was used for more reliable internet access during testing.

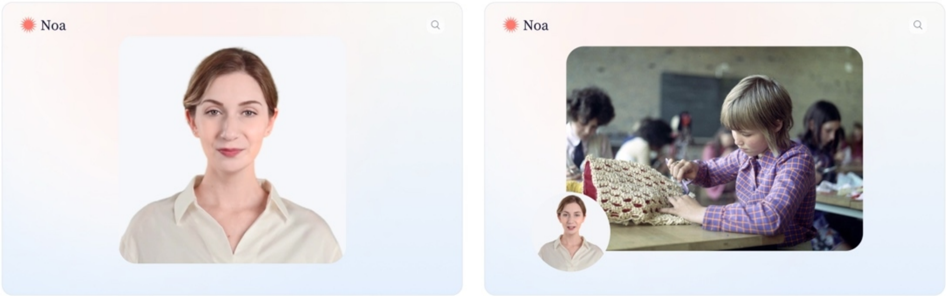

Figure 2: Stills of the UI of the Noa prototype. Left: the main screen. Right: the photo display for the RT flow.

Software

Noa was developed as a Vue-based web application, using Vuetify for a responsive and accessible user interface. The chatbot functionality was powered by OpenAI's AI models (accessed via their platform API), integrating the following components:

- whisper-1 for speech-to-text (STT) conversion;
- gpt-4o-2024-08-06 with JSON-based structured responses, enabling predefined actions such as displaying photos, responding to user input, and ending conversations;
- tts-1 for text-to-speech (TTS) conversion.

The first prototype relied on wake-word activation ("Hey Noa") via Google's SpeechRecognition API, but performance limitations within Chrome required adjustments. To improve accessibility, an on-screen activation button was added. In a second iteration, a key-binding feature enabled activation via a physical button on a tabletop device.

Personalisation features included user-specific prompts and dynamic avatar states, such as waiting and speaking modes. In the first prototype, users were alternately assigned either a human avatar or an animated otter. The second iteration introduced an animated abstract avatar composed of pulsating circles and droplets, enhancing state transitions and incorporating a processing mode.

USER TESTING AND RESULTS

User testing was conducted at two long-term care facilities specialising in dementia care. The first facility housed older adults, including individuals with moderate to severe dementia in psychogeriatric (PG) units, while the second specialised in younger persons (30–65) with early-onset dementia.

At the first facility, six participants (aged 82 to 95) tested Noa with either a human avatar or a cartoon-like otter. Some interacted through a Wizard of Oz setup, where a researcher simulated responses to explore engagement without the limitations of the early-stage AI system. At the second facility, three participants (aged 52 to 61) tested an abstract avatar chatbot and a version prompting discussions on daily challenges. The first test was conducted on a laptop due to time constraints, while the second used the hardware configuration we developed. Internet instability and microphone sensitivity posed challenges.

Engagement and Interaction

Engagement varied based on cognitive ability, topic familiarity, and tolerance for stimulation. Some participants engaged enthusiastically, particularly when discussing personal interests, while others became uncomfortable when Noa misinterpreted responses or interactions became overstimulating. One participant with aphasia (common in advancing dementia) struggled to engage, highlighting the current limitations of the system for users of different communication abilities.

Neither voice activation nor button-based interaction was successfully used without assistance. Caregivers or researchers had to intervene, indicating the need for more intuitive interaction mechanisms.

Participants involved in Wizard of Oz interactions were generally more positive compared to those interacting with the software prototype, with these participants responding more naturally, and reporting greater satisfaction with the interaction afterwards.

Avatar Perception

Avatar perception was highly individualised. Some responded well to the human avatar, while others rejected the otter (one participant resisted speaking with the animated otter avatar, stating "*I don't own a monkey, and I never will!*"), while other participants strongly preferred the otter. At the second facility, an abstract avatar was introduced to assess whether reducing anthropomorphic cues influenced interaction. Participants did not express a clear preference.

One participant in the first session continued engaging with Noa through voice interaction alone for several minutes after the screen of the prototype went black due to a technical malfunction, while another in the second session used the abstract animation to check whether Noa was "thinking" (processing input). This suggests that some users perceived the avatar as a social presence, while others relied on it for functional feedback.

Caregiver Perspectives

Caregivers saw potential for Noa in providing companionship during quiet moments and responding to repetitive questions to reduce workload. However, they noted that existing tools such as reminder watches and calendars were already effective for some users. They emphasised that Noa should focus on meaningful social interaction and support rather than duplicating existing functionalities.

At the second facility, caregivers highlighted the importance of Noa integrating into active daily routines of PwD rather than being stationary. Some participants preferred a portable version to support independence and flexible use.

Technical Challenges and Adaptability

On several occasions, microphone sensitivity and unreliable internet affected speech recognition and conversation flow. Noa sometimes failed to recognise the emotional valence of statements, leading to reactions that were occasionally unsettling for participants (e.g. a participant remarked about placing flowers on her mother's grave, and Noa responded that flowers can really cheer one up). It also frequently interrupted users or asked too many follow-up questions, which was technically difficult to control. These issues sometimes caused frustration, highlighting the need for better dialogue pacing and adaptive responses.

CONCLUSION

This study explored the feasibility of an ECA for delivering a simplified form of RT in long-term care, by exploring three hypotheses:

Hypothesis A: User testing confirmed that PwD could engage in basic conversations, though comprehension varied due to speech difficulties and instances where the system failed to accurately process spoken input, leading to misinterpretations, interruptions and response delays. Conversations flowed more smoothly in Wizard of Oz testing, highlighting the need for better AI response handling and contextual awareness to detect and respond to signs of distress or disengagement.

Hypothesis B: Avatar preferences were mixed, with no clear preference for human-like or abstract designs. In one instance, the animated otter was very poorly received, underscoring the need for thoughtful avatar selection. Voice-based interaction was favored, but wake-word activation didn't work without assistance.

Hypothesis C: Button activation proved more far accessible than voice-based activation, suggesting the need for multimodal options.

While this study focused on RT, caregivers saw potential for support during moments of confusion, particularly at night. They emphasised integrating Noa into existing routines rather than replacing effective tools. Technical challenges, particularly in speech recognition and conversation fluidity, highlight the need for further refinement.

DISCUSSION AND FUTURE RESEARCH

Several factors limited the effectiveness of Noa in facilitating independent interaction. Technical shortcomings, particularly in speech recognition and system responsiveness, required frequent caregiver intervention, reducing user engagement and autonomy. The strong individual variation in avatar perception suggests that a one-size-fits-all approach is likely insufficient, and personalisation should be central to future design iterations.

Emotional sensitivity remains a key challenge, as Noa was unable to detect or respond to distress, leading to moments of discomfort. Future research should develop more adaptive conversation flows that recognise signs of disengagement or distress, such as long pauses or repeated negative responses, to adjust interactions accordingly. Additionally, the testing environment influenced outcomes, as interactions were shaped by caregiver facilitation. Further studies should follow more structured protocols for caregiver instruction and explore how ECAs function in less structured, real-world care settings over extended periods.

While this study focused on RT, future research could examine Noa's potential in other forms of cognitive and emotional support, including passive engagement strategies for users who may not benefit from direct conversation. The identified need for portability and practical support for people with early-onset dementia raises an open question about how ECAs could enhance task execution and potentially radically improve technology accessibility for this group.

ACKNOWLEDGMENT

The authors would like to acknowledge Max Mikkers for his efforts in building the prototypes, and Alet van den Aker and Claudia van Gils from *Zorggroep Elde Maasduinen*, and Claudia Martens and Ted van der Zanden from *De Zorgboog* for their help in organising the on-site tests.

REFERENCES

Kelley, J. F. (1984) An iterative design methodology for user-friendly natural language office information applications. ACM Transactions on Information Systems, 2(1), pp. 26–41. doi: 10.1145/357417.357420.

Ministry of Health, Welfare and Sport (2020) National Dementia Strategy 2021–2030. The Hague: Ministry of Health, Welfare and Sport. Available at: https://www.government.nl/ministries/ministry-of-health-welfare-and-sport (Accessed: 27-02-2025).

Morales-de-Jesús, V., Gómez-Adorno, H., Somodevilla-García, M., & Vilariño, D. (2021). Conversational System as an Assistant Tool in Reminiscence Therapy for People with Early-Stage Alzheimer's. Healthcare, 9(8), 1036. doi: 10.3390/healthcare9081036.

Nielsen, J. (1993) Usability Engineering. San Diego: Morgan Kaufmann.

Norman, D. A. (2013) The Design of Everyday Things. Revised and expanded edition. New York: Basic Books.

OpenAI (2024) Introducing structured outputs in the API. Available at: https://openai.com/index/introducing-structured-outputs-in-the-api/ (Accessed: 27-02-2025).

Rampioni, M., Stara, V., Felici, E., Rossi, L., & Paolini, S. (2021). Embodied Conversational Agents for Patients With Dementia: Thematic Literature Analysis. JMIR Mhealth Uhealth, 9(7), e25381. doi: 10.2196/25381.

Smith, J., Doe, A., Brown, C. & Johnson, R. (2023) Demonstration of the AMPER System for Individuals with Alzheimer's Disease. Proceedings of the International Conference on AI and Dementia Studies, Edinburgh, UK, 15–17 July 2023. doi: 10.1145/3652988.3696201.

Woods, B., O'Philbin, L., Farrell, E. M., Spector, A. E., & Orrell, M. (2018) Reminiscence therapy for dementia. Cochrane Database of Systematic Reviews, Issue 3, CD001120. doi: 10.1002/14651858.CD001120.pub3.