From User Feedback to Home Deployment: Designing a Virtual Human App for Older Adults with Mild Cognitive Impairment

Y. Gao, N. Kerse, D. Orbaugh, and E. Broadbent

Abstract— **Virtual humans may provide scalable support for cognitive health in later life, but many existing systems for older adults rely on scripted interaction or have been evaluated only in brief laboratory settings. This paper presents the design of a large language model (LLM)-enabled virtual human app for older adults with mild cognitive impairment (MCI) and outlines a planned home-based feasibility evaluation. The app centers on Mary, a virtual human character embedded across the interface, and supports three activity domains: guided physical exercises, cognitive games, and social conversation. The design was informed by an earlier user-centered study in which older adults with memory concerns and aged-care experts interacted with two social robots and a virtual human. Their feedback guided the choice of a virtual human app and informed key design features, including conversation memory, real-time exercise feedback, and progress-based rewards. The planned study will use a single-arm mixed-methods feasibility design in which 15 older adults with MCI will use the app at home for one week. Primary outcomes will be the feasibility and acceptability of the app and study procedures. This work contributes an example of translating user feedback into a home-deployable virtual human intervention for cognitive health support.**

I. Introduction

Mild cognitive impairment (MCI) is characterized by cognitive decline beyond what would be expected for a person's age and education level, but that does not significantly affect daily functioning. People with MCI are at increased risk of progression to dementia, where cognitive impairment becomes more severe and significantly interferes with everyday functioning [1]. Although pharmacological treatments for cognitive impairment remain limited, health behaviors such as regular physical exercise, cognitive activity, and social interaction have been shown to reduce the risk of dementia and support cognitive health [2]. There is therefore a need for accessible interventions that help older adults maintain these activities in daily life.

Social technologies, such as social robots and virtual humans, may provide a means to support these behaviors. Social robots have been used to prompt and guide older adults through physical exercises, cognitive games, and conversations [3]. Virtual humans have similarly been explored in geriatric care for companionship, health instruction, and activity guidance [4]. Recent advances in large language models (LLMs) may allow these technologies to provide more natural, flexible, and personalized conversations [5]. However, LLMs also raise ethical and design concerns, such as privacy risks, inappropriate advice, and potential overreliance [6].

To examine how these opportunities and concerns apply to older adults with MCI, we developed a three-study project to design and evaluate social technologies for cognitive health. Study 1 explored the views of older adults with memory concerns, including MCI and subjective memory complaints, as well as aged-care experts. Participants interacted with a humanoid robot (Pepper, Aldebaran Robotics), a pet-like robot (Paro, AIST), and a virtual human (Mary, developed using Soul Machines). Both Pepper and Mary incorporated LLM-enabled conversation. Feedback suggested that LLM-enabled conversation was acceptable, but participants wanted the technology to remember previous conversations. Participants also emphasized the value of real-time feedback during exercises and progress-based rewards to support motivation. Additionally, they highlighted the need for transparency about the system's artificial nature and clear limits around advice-giving. These findings informed the current study by supporting the inclusion of conversation memory, real-time exercise feedback, progress summaries and rewards, while limiting the LLM to low-risk social conversation and activity encouragement.

Study 1 also informed the decision to focus the current study on a virtual human app rather than a physical robot. Although some participants found physical robots engaging, robots introduced practical challenges for home deployment, including cost, transport, maintenance, charging, space requirements, and technical reliability. A virtual human app can be delivered through familiar devices such as smartphones, tablets, or computers, making it more scalable and easier to integrate into everyday home settings.

This paper presents Study 2 of the project. We first review related work on virtual human interventions for older adults and identify gaps in the literature. We then describe the design of the app and outline a planned one-week home feasibility evaluation.

II. Related Work

A. Virtual Humans for Older Adults

Virtual humans are screen-based, human-like characters that can communicate with users through speech, gesture, movement, and facial expression [7]. A recent integrative review by Dino et al. [4] examined 36 studies of virtual humans in geriatric care. The review found that they have mainly been used for social health interventions, particularly

Y. Gao is with the Department of Psychological Medicine, University of Auckland, New Zealand (corresponding author, e-mail: yuan.gao@auckland.ac.nz).

N. Kerse is with the Department of General Practice and Primary Healthcare, University of Auckland, New Zealand (e-mail: n.kerse@auckland.ac.nz).

D. Orbaugh is with the Bioengineering Institute, University of Auckland, New Zealand (e-mail: derek.orbaugh@auckland.ac.nz).

E. Broadbent is with the Department of Psychological Medicine, University of Auckland, New Zealand (e-mail: e.broadbent@auckland.ac.nz).

Manuscript 687 submitted to 2026 IEEE 35th International Conference on Robot and Human Interactive Communication (RO-MAN). Received June 12, 2026.

companionship, although the field is beginning to shift toward educational and health-related support. For example, virtual humans have been used to provide health information, guide users through assessments, support daily routines, deliver reminders, encourage self-management, and assist with activities such as relaxation or exercise guidance [4]. However, most systems used pre-programmed responses, and the use of artificial intelligence in virtual human development remains limited. This suggests a gap for more flexible and adaptive virtual human interventions for older adults, particularly systems designed for sustained use.

More relevant to our study, Zsoldos et al. [8] evaluated a virtual human designed to improve engagement in computerized cognitive training. The study recruited 104 participants, including 52 older adults aged 60-79 and 52 young adults. Participants completed one laboratory-based cognitive training session with four cognitive exercises, supervised by a virtual assistant animated through a Wizard-of-Oz method. The virtual assistant was perceived as useful, and older adults rated the assistant and the cognitive training more positively than younger adults. However, the study was a single-session laboratory study with healthy older adults, and the authors noted that future work should test the system with people experiencing MCI and evaluate longer-term use.

There is limited evidence on virtual humans embedded within health apps for older adults. One relevant example is the study by ter Stal et al. [9], who implemented a virtual human within a frailty assessment app. In the app, three health questionnaires were translated into agent-led dialogues. The study used a within-subject design with 20 older adults, comparing a young female agent and an older male agent. The app received high usability and intention-to-use scores, but participants did not perceive added value from the agent itself. The authors suggested that future versions may need richer interaction features, such as speech output. Participants also expressed interest in receiving advice on actions they could take to improve their health, such as showing exercise videos to support physical strength.

B. *The Present Study*

In summary, three gaps in the existing literature motivate the present study. First, many virtual human systems for older adults rely on pre-programmed responses, which may limit personalization and sustained engagement. Second, relatively little work has focused specifically on older adults with MCI, who may have specific needs. Third, there is limited evidence on virtual humans embedded within app-based interventions for repeated home use. The present study addresses these gaps by developing an LLM-enabled virtual human app for older adults with MCI and evaluating its feasibility and acceptability during one week of home use.

III. APP DESIGN AND DEVELOPMENT

The prototype has been developed as a home-based cognitive health app centered on a virtual human character named Mary. The name was selected based on Study 1 feedback from older adults with memory concerns, who perceived it as familiar and easy to remember. In the current prototype, Mary functions as a static virtual human character that supports speech input and output rather than a fully animated agent. She appears across screens to provide a recognizable social identity for the app. This design reduces technical complexity for the initial home feasibility study. Participants' feedback will inform whether animation should be added in Study 3.

The app includes three activity modules: physical exercises, cognitive activities, and conversation. The activities within each module are consistent with those explored in Study 1, where participants perceived them as acceptable and useful. Across the app, screens use large text, large buttons, and a limited number of choices. This is intended to make each step visible and predictable, reducing cognitive and navigational demands. The app also includes accessibility controls, such as options to repeat instructions or slow down speech, and status cues such as "Mary is speaking" and "Mary is listening" to support turn-taking.

A. *Physical Exercise Module*

The physical exercise module guides users through three lower-limb movements: knee bends, sit-to-stand, and side hip strengthening. These movements were selected from the Otago Exercise Programme, which has been widely validated and shown to reduce mortality and falls in older adults [10]. The selected movements are low intensity and can be completed at home with minimal equipment.

Each exercise follows a sequence of preparation, demonstration, practice, and feedback. For example, when users select sit-to-stand, Mary provides text- and image-based instructions on how to position the device, ensure the whole body is visible on screen, use a stable chair, and complete the expected number of repetitions and sets. Users can also choose to watch a video demonstration before starting the exercise. During the exercise, the app displays exercise progress, including the number of completed and remaining repetitions.

A key feature of this module is real-time exercise feedback. Using the device camera, the app detects the user's movement and provides corrective prompts, such as asking the user to move more slowly or adjust their posture. Video footage is processed only to generate feedback and will not be recorded or stored.

After the exercise, users receive a summary showing the number of sets completed, total repetitions, stars earned, and daily progress. If the activity is not completed, Mary provides nonjudgmental feedback and allows users to try again, choose another exercise, or return to the home screen.

B. *Cognitive Activity Module*

The cognitive activity module includes two games: a category game and a word association game. These games were chosen because they are simple to understand, use familiar word-based tasks, and can be completed by typing or speaking. The aim is to provide brief cognitive stimulation without requiring users to learn complex rules.

In the category game, users are asked to name as many items as they can from a given category within two minutes. The screen shows the category, the time remaining, the answers already given, and a box where users can type or say their next answer. In the word association game, users are given a prompt word and asked to say or type related words.

For example, if the prompt is “beach,” they might respond with words such as “sand,” “water,” or “holiday.”

Mary supports the games by introducing each task, providing encouragement, and offering optional help. Users can ask for a clue, request repeated instructions, or ask Mary to speak more slowly.

C. Social Conversation Module

The conversation module allows users to chat with Mary by speaking or typing. Users can start a conversation freely, or they can choose from suggested prompts, such as talking about their day, discussing hobbies, or asking Mary to tell a joke. These prompts are included to reduce the effort needed to start a conversation.

The module uses a GPT-5.2-based dialogue system to support flexible social conversation. Conversation memory allows Mary to refer to previous topics, helping interactions feel more continuous and personally relevant during repeated use. Because participants may have memory difficulties, the app will not rely on a one-off reminder about privacy and boundaries of advice. A brief safety message will remain visible on the conversation screen, reminding users that Mary is an artificial system, that she cannot provide diagnosis, treatment advice, or personalized medical guidance, and that users should avoid entering sensitive personal or medical information. The dialogue system also implements predefined safety instructions that direct Mary to acknowledge requests involving health or safety concerns and, where appropriate, encourage users to seek appropriate human support.

IV. PLANNED FEASIBILITY EVALUATION

Recruitment for the study is currently being prepared. The planned study will use a single-arm mixed-methods feasibility design to evaluate one-week home use of the virtual human app by older adults with MCI. The aim is to examine whether the app and study procedures are feasible and acceptable.

A. Participants

The planned sample will include 15 older adults with MCI. The sample size is appropriate for a feasibility study because the primary aim is to identify usability issues, implementation barriers, and participant feedback rather than to test effectiveness. Participants will be eligible if they are aged 65 years or older, able to provide informed consent, and able to communicate in English. Individuals with diagnosed dementia or severe sensory, psychiatric, neurological, or motor difficulties that would prevent meaningful app use will be excluded.

B. Procedure

Fig. 1 summarizes the planned study procedure. Participants will be encouraged to use the app at home for at least 15 minutes per day for a week. With permission, app usage data will be collected to describe engagement across activities. The postintervention interview will provide qualitative insight into participants’ experiences, including perceived benefits, difficulties, and suggestions for refinement. A two-week follow-up will explore whether any behaviors introduced during the study are maintained after app access is removed, and whether discontinuation leads to any negative experiences, such as missing the virtual human app or reduced motivation to continue activities.

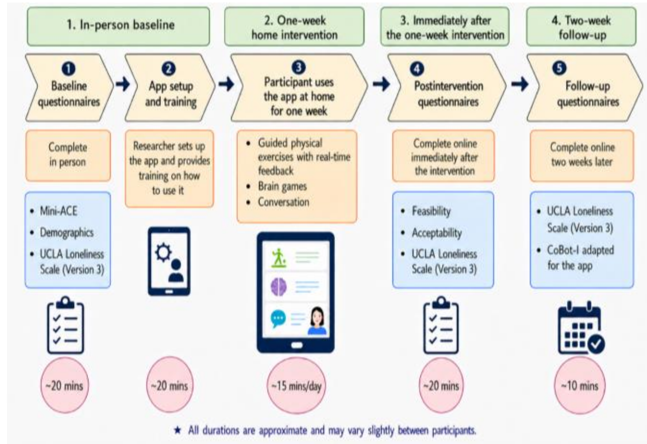

Fig. 1. An overview of the planned study procedure. Mini-ACE: Mini-Addenbrooke’s Cognitive Examination [11]; CoBot-I: Companion Robot Impact Scale.

C. Outcomes

The primary outcomes will be the feasibility and acceptability of the app and the study procedures. App feasibility will be assessed by recording whether participants can access the app activities from home, the number and type of technical support contacts required, and whether technical issues interfere with activity completion. Feasibility of the study procedures will be assessed through recruitment and retention rates, reasons for withdrawal, missing data, incomplete questionnaires, and any difficulties participants experience during the study.

Acceptability of the app will be assessed using self-report ratings, behavioral engagement data, and qualitative feedback. Participants will rate their satisfaction with the app activities, the likelihood that they would continue similar activities in the future, their willingness to use the app again, and their comfort using the app independently at home. They will also be asked whether they would recommend the app to someone they care about. Acceptability of the study procedures will be explored through open-ended questions.

Engagement will be examined using system log data. Usage data will include frequency of use, duration of daily use, and activity completion across the modules. These data will help identify which parts of the app are most used and where refinements may be needed. Conversation content will not be collected.

Secondary outcomes will be examined descriptively and exploratorily. Loneliness will be assessed at baseline, postintervention, and two-week follow-up using the 20-item UCLA Loneliness Scale (Version 3) [12] to examine the direction and pattern of change over time. A modified version of the Companion Robot Impact Scale developed by Doraiswamy and colleagues (2022) will also be used to explore participants’ perceptions of the virtual human app’s impact on companionship, social connection, loneliness, mood, mental stimulation, and everyday wellbeing. Findings will be used to guide future refinement and trial planning.

D. Ethical Considerations

Attention will be given to participants' views on conversation memory. Although memory may make interactions feel more personalized and continuous, it also raises questions about what information is remembered, how it is used, and how much control users should have over it. The postintervention interview will explore participants' comfort with remembered information, privacy expectations, data storage concerns, and preferences for managing or deleting personal information in future versions of the app.

V. Expected contributions

This work contributes to the RO-MAN community in three ways. First, it presents an LLM-enabled virtual human app specifically designed for older adults with MCI, a group that has received limited attention in prior virtual human research. The prototype integrates physical exercise, cognitive activities, and social conversation within one home-use system, rather than focusing on a single activity domain. Second, the work contributes a practical example of translating earlier user-centered findings into a deployable prototype. Study 1 feedback informed the choice of platform and the app's design features, which shows how user-centered findings can be operationalized in a prototype. Third, the planned study will provide early evidence on whether this type of app can be used independently in older adults' homes. This is important because much previous work has evaluated virtual humans in brief interactions, laboratory settings, or scripted systems.

VI. Future work

The planned feasibility evaluation will inform Study 3, which will refine the app and prepare it for a larger evaluation. Findings from Study 2 will be used to identify which features are acceptable, which require modification, and what additional support older adults with MCI may need during home use. The larger evaluation may involve a longer home-use period, a larger sample, and a more rigorous study design, such as a pilot randomized controlled trial comparing the app with usual activities or an active control condition. This would allow future research to examine preliminary effects on cognitive health, loneliness, and perceived companionship.

Participants' experiences of Mary as a virtual human character will also inform future development. In the current prototype, Mary is static and provides social presence through visual identity and speech-based interaction. If participants indicate that movement, facial expression, or gesture would make Mary feel more engaging, understandable, or supportive without increasing cognitive or technical burden, a more animated virtual human will be considered for the next phase.

References

[1] S. Gauthier *et al.*, "Mild cognitive impairment," *The Lancet*, vol. 367, no. 9518, pp. 1262–1270, Apr. 2006.

[2] G. Livingston *et al.*, "Dementia prevention, intervention, and care: 2024 Report of the Lancet Standing Commission," *The Lancet*, vol. 404, no. 10452, Jul. 2024, doi: https://doi.org/10.1016/s0140-6736(24)01296-0.

[3] R. C. Preston, M. R. Shippy, C. M. Aldwin, and N. T. Fitter, "How can robots facilitate physical, cognitive, and social engagement in skilled nursing facilities?," *Frontiers in Aging*, vol. 5, Nov. 2024, doi: https://doi.org/10.3389/fragi.2024.1463460.

[4] M. J. S. Dino *et al.*, "Virtual Humans in Geriatric Care: An Integrative Review," *The Journals of Gerontology, Series A: Biological Sciences and Medical Sciences*, vol. 80, no. 9, Jul. 2025, doi: https://doi.org/10.1093/gerona/glaf145.

[5] M. Pinto-Bernal, M. Biondina, and T. Belpaeme, "Designing Social Robots with LLMs for Engaging Human Interaction," *Applied Sciences*, vol. 15, no. 11, p. 6377, Jun. 2025, doi: https://doi.org/10.3390/app15116377.

[6] J. Haltaufderheide and R. Ranisch, "The ethics of ChatGPT in medicine and healthcare: a systematic review on Large Language Models (LLMs)," npj digital medicine, vol. 7, no. 1, Jul. 2024, doi: https://doi.org/10.1038/s41746-024-01157-x.

[7] D. Burden and M. Savin-Baden, Virtual humans: Today and tomorrow. Chapman and Hall/CRC, 2019.

[8] I. Zsoldos et al., "The Value of a Virtual Assistant to Improve Engagement in Computerized Cognitive Training at Home: Exploratory Study," JMIR Rehabilitation and Assistive Technologies, vol. 11, p. e48129, Jun. 2024, doi: https://doi.org/10.2196/48129.

[9] S. ter Stal, M. Broekhuis, L. van Velsen, H. Hermens, and M. Tabak, "Embodied Conversational Agent Appearance for Health Assessment of Older Adults: Explorative Study," JMIR Human Factors, vol. 7, no. 3, p. e19987, Sep. 2020, doi: https://doi.org/10.2196/19987.

[10] Z. He et al., "The effectiveness of digital technology-based Otago Exercise Program on balance ability, muscle strength and fall efficacy in the elderly: a systematic review and meta-analysis," BMC Public Health, vol. 25, no. 1, Jan. 2025, doi: https://doi.org/10.1186/s12889-024-21251-9.

[11] A. Grasina, H. Espirito-Santo, L. Lemos, Maria Manuela Vilar, Luís Simões-Cunha, and F. Daniel, "Mini-ACE: Validation Study Among Older People in Long-Term Care," Journal of cognition, vol. 7, no. 1, pp. 5–5, Jan. 2024, doi: https://doi.org/10.5334/joc.330.

[12] D. W. Russell, "UCLA Loneliness Scale (Version 3): Reliability, Validity, and Factor Structure," *Journal of Personality Assessment*, vol. 66, no. 1, pp. 20–40, 1996, doi: https://doi.org/10.1207/s15327752jpa6601_2.

Manuscript 687 submitted to 2026 IEEE 35th International Conference on Robot and Human Interactive Communication (RO-MAN). Received June 12, 2026.