ReMe: Scaffolding Personalized Cognitive Training via Controllable LLM-Mediated Conversations

Zilong Wang*
Microsoft Research
Shanghai, China
wangzilong@microsoft.com

Nan Chen*
Microsoft Research
Shanghai, China
nanchen@microsoft.com

Luna K. Qiu
Microsoft Research
Shanghai, China
lunaqiu@microsoft.com

Ling Yue
Department of Geriatric Psychiatry, Shanghai Mental Health Center, Shanghai Jiao Tong University School of Medicine
Shanghai, China
bellinthemoon@sjtu.edu.cn

Geli Guo
Microsoft Research
Beijing, China
v-guobella@microsoft.com

Yang Ou
Microsoft Research
Beijing, China
yang.ou@microsoft.com

Shiqi Jiang
Microsoft Research
Shanghai, China
shijiang@microsoft.com

Yuqing Yang
Microsoft Research
Shanghai, China
Yuqing.Yang@microsoft.com

Lili Qiu
Microsoft Research
Shanghai, China
liliqiu@microsoft.com

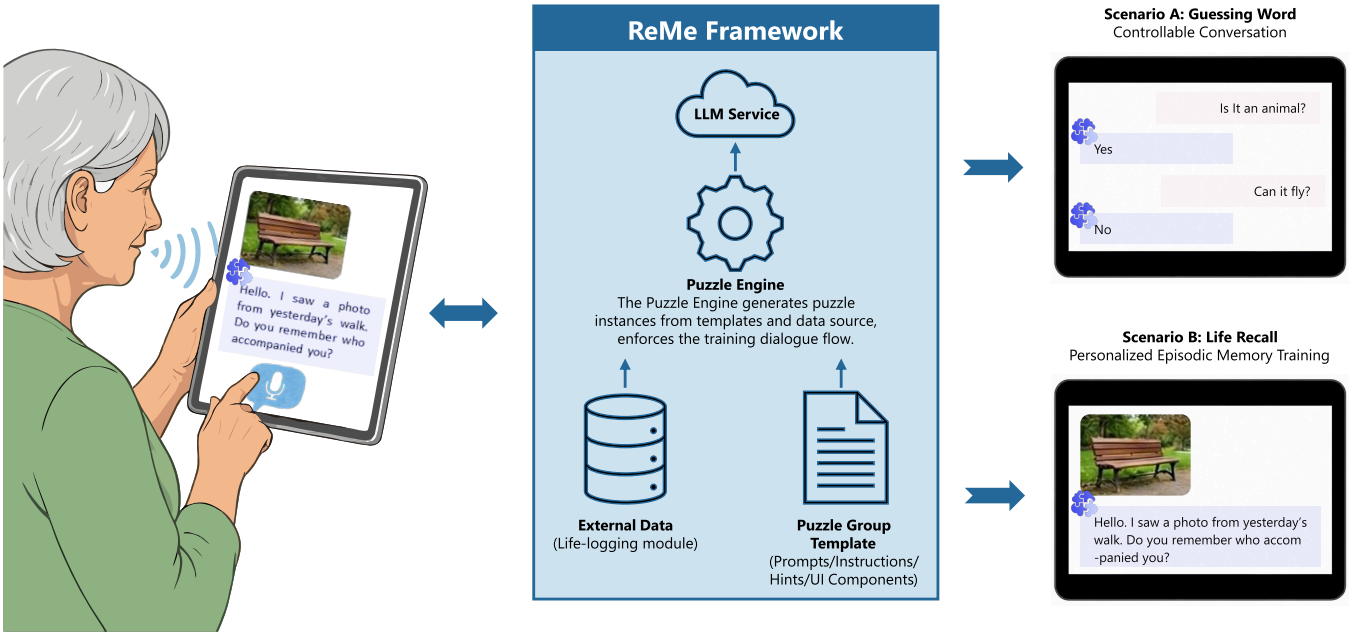

Figure 1: ReMe overview: Users interact with an LLM-powered voice-based chatbot through a multimodal training interface. The Puzzle Engine integrates information from life logs (and other data sources when needed) to create a puzzle instance and manage the conversational workflow.

Abstract

Global aging calls for scalable and engaging cognitive interventions. Computerized cognitive training (CCT) is a promising non-pharmacological approach, yet many unsupervised programs rely

*These authors contributed equally to this work.

CHI EA '26, Barcelona, Spain

ACM ISBN 979-8-4007-2281-3/2026/04
https://doi.org/10.1145/3772363.3798695

on rigid, hand-authored puzzles that are difficult to personalize and can hinder adherence. Large language models (LLMs) offer more natural interaction, but their open-ended generation complicates the controlled task structure required for cognitive training. We present ReMe, a web-based framework that scaffolds cognitive training through controllable LLM-mediated conversations, addressing both rigidity in conventional CCT content and the need for conversational controllability. ReMe features a modular Puzzle Engine that represents training activities as reusable puzzle groups specified by structured templates and constraint rules, enabling rapid development of dialogue-based word games and personalized tasks grounded in user context. By integrating personal life logs, ReMe supports Life Recall activities for episodic-memory practice through guided retrieval and progressive cues. A community pilot with 32 adults aged 50+ provides initial feasibility signals.

CCS Concepts

• **Human-centered computing** → **Empirical studies in HCI**; *Interactive systems and tools*; • **Applied computing** → *Health care information systems.*

Keywords

AI chatbot, cognitive training, LLM, digital health

ACM Reference Format:

Zilong Wang, Nan Chen, Luna K. Qiu, Ling Yue, Geli Guo, Yang Ou, Shiqi Jiang, Yuqing Yang, and Lili Qiu. 2026. ReMe: Scaffolding Personalized Cognitive Training via Controllable LLM-Mediated Conversations. In *Extended Abstracts of the 2026 CHI Conference on Human Factors in Computing Systems (CHI EA '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3772363.3798695

1 Introduction

The proportion of older adults has increased substantially over recent decades [17], accompanied by a growing burden of cognitive disorders such as Alzheimer's disease [4]. Currently, no treatment can reliably reverse Alzheimer's [28]; therefore, prevention and early intervention remain essential. In this context, cognitive training has emerged as a promising non-pharmacological strategy.

Computerized cognitive training (CCT) can improve certain cognitive functions in older adults [9, 13], including long-term benefits for domains such as reasoning and processing speed observed in the ACTIVE trial [19, 26]. However, real-world impact in unsupervised settings remains constrained by a design gap. Many scalable programs rely on rigid, hand-authored puzzles to preserve task structure, making content difficult to personalize. This rigidity can undermine engagement and adherence [15, 23]. Moreover, effective interventions often still depend on professional supervision for sustained use and usability support, while unsupervised training tends to show smaller effects [10]. This is particularly challenging for episodic-memory practice, where ecologically grounded tasks are difficult to deliver without supervision [18, 20, 33].

LLM-powered voice-based chatbots create new opportunities for cognitive training. Compared with traditional interfaces, conversational interaction offers a low-barrier way to engage older adults, while LLMs provide open-world knowledge that can support more diverse tasks and enable personalization using user profiles, performance history, and everyday records [2, 8]. LLM-based chatbots have also shown promise in medicine and mental health support, motivating careful exploration in cognitive health contexts [1, 7, 14]. However, cognitive training relies on structured tasks, enforceable interaction constraints, and interpretable feedback to ensure consistent practice and track progress. Open-ended generation can drift off-task or break required formats, weakening training fidelity and progression. This creates a core tension for LLM-mediated cognitive training: how to preserve natural, engaging dialogue while maintaining the controllability needed for structured practice. ReMe addresses this tension by scaffolding cognitive training through controllable LLM-mediated conversations.

We present **ReMe**, a web-based framework for building LLM-powered chatbots for personalized cognitive training. ReMe is guided by four design goals: **(DG1) controllable dialogue-based training** that preserves task objectives while leveraging conversational engagement, drawing on scaffolding principles to maintain focus and reduce degrees of freedom [11, 27]; **(DG2) personalized episodic-memory tasks** grounded in each user's life context for unsupervised practice, informed by the self-reference effect [22] and utilizing life logs as autobiographical memory cues [21] (rather than "verifiable facts"); **(DG3) low-friction interaction for older adults** through voice-first and lightweight UI to accommodate age-related declines in fine motor skills [12] and minimize extraneous cognitive load [16] via natural modalities [24]; and **(DG4) an extensible puzzle framework** for rapid creation and iteration to simplify LLM prompting complexities [30] and sustain adherence through task variety. Concretely, ReMe represents training activities as reusable *puzzle groups* with explicit templates (prompts, instructions, hints, and interaction components) and enforceable constraint rules, and connects them with **personal life logs** that provide personally grounded real-life cues for guided episodic-memory practice (Fig. 1).

Contributions:

- **ReMe**, a reusable framework for rapidly prototyping LLM-mediated cognitive training with modular task abstractions and reusable interaction components.
- A controlled-generation interaction paradigm that supports training-oriented conversations through enforceable constraints and structured task progression.
- Feasibility evidence from a public outreach pilot study with older adults (N=32, aged 50+) showing initial usability and engagement signals.

2 System Design and Implementation

ReMe centers around three components (Fig. 1). Guided by the design goals introduced above, the **Puzzle Engine** instantiates training tasks from reusable specifications and manages task-oriented conversational workflows. The **Life-Logging Module** supplies personal episodic materials for scenario-based memory training. The **Training User Interface** supports voice-first multimodal dialogue with reusable interaction components beyond conventional chat.

2.1 Puzzle Engine: Puzzle Groups and Instance Creation

ReMe organizes tasks into *puzzle groups*, which are reusable task definitions that share objectives, instructions, and interaction rules. Each puzzle group includes a group name, a prompt template used for model inference, an instruction template shown to users, a hint template describing how hints are delivered (including multimodal hints when needed), and optional interaction components required by the task (e.g., rating, end-of-session, drawing). Puzzle groups also specify explicit interaction constraints, such as required response formats (e.g., yes/no-only answering), to preserve training fidelity (DG1). At runtime, the engine creates a puzzle instance by populating templates with relevant data sourced from randomized generation, external sources, or life logs, prepares any hint payloads (e.g., a related life-log photo), and initiates a chat session. This template-based authoring enables rapid customization and iteration without implementing a new application for each task (DG4).

To balance natural dialogue with training controllability (DG1), ReMe encodes task constraints and recovery behaviors in the puzzle-group system prompt. The model is instructed to perform internal planning before producing a user-visible response, and to steer the dialogue back to valid interaction forms when violations occur. Currently, constraint enforcement is prompt-level: the system prompt specifies valid interaction patterns per puzzle group, and the model performs a ReAct-style reasoning step over the full session history each turn to detect violations and issue redirects before advancing the task. [29] For example, in a yes/no riddle task, if a user requests disallowed information (e.g., "What is the first letter?"), the chatbot restates the constraint and redirects the user to a valid yes/no question about properties or usage.

2.2 Life-Logging Module

For episodic-memory training, ReMe includes a life-logging module where users can upload life log entries containing text and images. The module stores timestamps, descriptions, and images, and supports later retrieval for training. These life logs provide personally grounded content that can be transformed into individualized recall tasks (e.g., recalling a meal or an event from the previous day), enabling personalization in unsupervised practice (DG2). The retrieved logs also support progressive hinting with textual or image cues, allowing the system to provide assistance while keeping tasks grounded in real-life materials (DG2). To mitigate hallucination, the system prompt restricts factual claims to the retrieved log entry and instructs the model to explicitly ground each verification step in the retrieved artifacts before producing a response [31].

2.3 Training User Interface and Reusable Components

ReMe defines *Chat Message* as the basic unit of conversation, containing voice, text, and images. Messages are model-independent and can be converted into the input format required by a chosen model, keeping puzzle definitions decoupled from any specific LLM (DG4). The UI supports voice-first interaction complemented by lightweight widgets to reduce operational burden for older adults (DG3). To support specialized puzzle interactions, ReMe provides reusable UI components invoked via tags and parameters in the chatbot output, such as: (i) *hint* to display pre-configured multimodal messages, (ii) *rating* to collect structured feedback, (iii) *end* to terminate a session, and (iv) *draw* to enable drawing input (e.g., a clock drawing task). These components externalize task control and structured input beyond free-form dialogue, making interactions consistent across puzzle types and easy to extend (DG4), while maintaining low-friction use in voice-based sessions (DG3).

(a) Guessing Word. **(b) Life Recall.**

Figure 2: Training cases instantiated by ReMe.

3 User Scenario

We illustrate ReMe with two puzzle groups that cover constrained dialogue for structured training and life-log grounded episodic recall.

3.1 Persona and Context

Consider an older adult user engaging in short daily training sessions at home, primarily through voice with lightweight UI support when needed (DG3).

3.2 Scenario A: Guessing Word (Open-World Yes/No Riddle)

Goal. Train reasoning and language through an open-world guessing game under a strict yes/no answering constraint (DG1). This stepwise, information-seeking interaction is intended to engage working memory, as the user must maintain and update partial hypotheses across turns while actively drawing on and refining everyday language skills and commonsense world knowledge.

Workflow. The user is given a category and asks a sequence of yes/no questions to narrow down a hidden target. The chatbot answers in yes/no form, enforces the format, and may summarize confirmed facts to guide subsequent questions.

Mechanism. Each session is defined by a category and a target word. The puzzle group encodes knowledge-based yes/no responding, violation steering for invalid queries, brief fact summarization, and acceptance of equivalent guesses.

Example. In Fig. 2a, the category is *vehicle* and the target is *bicycle*. After a valid question ("Is it in the sky?") receives "no," the user asks an invalid request ("What is the first letter?"). The chatbot restates the constraint, redirects the user to ask a valid

yes/no question, summarizes known facts, and accepts "bike" as equivalent to *bicycle*.

3.3 Scenario B: Life Recall (Episodic Memory)

Goal. Prompt users to recall details of a recent life event recorded in a life log, with progressive cues when needed (DG2). In doing so, the task supports individualized training of episodic memory grounded in the user's own autobiographical history.

Workflow. The user first records daily events as life logs. During training, the system retrieves a relevant entry and prompts the user to recall details. The chatbot verifies responses and provides progressively stronger hints, including an optional photo cue.

Mechanism. The puzzle instance is generated from a retrieved life-log entry. The chatbot guides recall with structured prompts, verifies recalled content against logged text and images, and adapts hint specificity based on user progress.

Example. In Fig. 2b, the system retrieves a logged "yesterday's lunch" entry and prompts recall. The chatbot confirms recalled items, provides increasingly specific textual hints when recall is incomplete, and can reveal the logged photo to support retrieval.

3.4 What the Scenarios Demonstrate

Both puzzles are authored as puzzle groups that reuse templates, constraints, and UI components, enabling rapid iteration without implementing task-specific interfaces from scratch (DG4).

4 Pilot Study: Usability and Engagement Signals

We conducted a community outreach pilot study to assess feasibility signals of ReMe in a low-support setting, focusing on usability and engagement rather than cognitive efficacy. The pilot was conducted on World Alzheimer's Day following a public awareness lecture. After a brief introduction, participants voluntarily tried both puzzles on designated devices and completed a questionnaire.

Ethics and privacy. The pilot study was approved by the Institutional Review Board (IRB). To minimize privacy risks, the Life Recall puzzle used a pre-set image as a case study rather than participants' personal life logs. As a result, the pilot is limited to evaluating the interaction flow and usability of Life Recall with generic content; testing with participants' own life-log data remains an important next step. No identifiable personal information was collected or retained, and all session data were cleared after the session concluded.

Participants. We report questionnaire results from 32 participants aged 50 years or older (n=32). Age distribution was: 50–60 (n=9, 28.13%), 60–70 (n=15, 46.88%), and 70+ (n=8, 25.00%). Gender distribution was: female (n=23, 71.88%) and male (n=9, 28.13%). Within the same analysis sample, 84.38% reported at least a high-school education, 37.50% reported taking daily medication for chronic conditions, and 75.00% reported using a mobile phone for more than 2 hours per day. In addition, 25.00% reported prior experience with cognitive training software, and half of these participants reported that the interfaces of previous tools were unfriendly.

Measures. Participants rated perceived difficulty and enjoyment for each puzzle using 5-point Likert scales (difficulty: 1=easiest, 5=hardest; enjoyment: 1=least enjoyable, 5=most enjoyable).

Results. Both puzzles were rated as moderately difficult and enjoyable. For *Guessing Word*, difficulty was median=3 (IQR=2–3) and enjoyment was median=3 (IQR=2–4); 81.25% rated difficulty as 2–4/5 and 71.88% rated enjoyment as ≥3/5. Participants explicitly highlighted the benefits of the voice-first interface compared to touching screens: "*This(ReMe) feels flexible because I can speak directly. Previous training tools required interaction with the phone, which was difficult for me; this is much easier.*"

For *Life Recall*, difficulty was median=3 (IQR=2–3) and enjoyment was median=3 (IQR=3–4); 81.25% rated difficulty as 2–4/5 and 75.00% rated enjoyment as ≥3/5. Although the pilot used generic images, participants expressed a strong desire for personalization: "*The recall task is interesting, it would be more engaging if the content were based on my own life experiences when training.*"

For longer-term willingness, 31.25% reported being willing to invest at least 30 minutes per day to reduce Alzheimer's risk, and 90.62% indicated they would likely continue training for at least one month if recommended.

5 Discussions, Limitations, and Risks

The pilot suggests that ReMe-based puzzles can be usable and engaging in a low-support setting. Together with our two case studies, these results indicate ReMe's potential to deliver both episodic-memory practice grounded in everyday records and open-world reasoning and language training under explicit interaction constraints. More rigorous studies are needed to evaluate cognitive outcomes.

Design implications. ReMe illustrates a controllable conversational training paradigm that preserves task fidelity through explicit constraints and recovery behaviors, while retaining low-barrier dialogue. It also demonstrates a path from life logs to ecologically grounded episodic-memory practice via retrieval-based prompting and progressive cues. As a reusable framework, ReMe lowers the cost of authoring and iterating training tasks through puzzle-group templates and reusable UI components.

Limitations. Our evidence is limited to short-term feasibility signals and does not establish efficacy or transfer. Voice-first, turn-based interaction may still feel less natural than human conversation and can be sensitive to noisy environments, requiring onboarding for some users. High-quality interaction depends on capable LLMs, introducing cost and latency. Personalized recall quality also depends on life-log coverage.

Risks and mitigation directions. Life-log based personalization raises privacy and security risks; mitigations include data minimization, user controls, and secure storage [3, 25]. LLM hallucinations may harm safety and training validity; mitigations include grounding in retrieved artifacts, conservative prompting, and output filtering [5, 6, 32]. Prolonged empathetic interaction may increase over-reliance; mitigations include transparency, pacing or usage limits, and integration with real-world support.

Future work. We will expand ReMe with additional puzzle groups and difficulty adaptation, and conduct longitudinal studies with validated cognitive assessments. We will further optimize end-to-end responsiveness for voice-first interaction and investigate privacy-preserving approaches for handling life-log data in real-world deployments.

References

[1] Alaa A Abd-Alrazaq, Mohannad Alajlani, Ali Abdallah Alalwan, Bridgette M Bewick, Peter Gardner, and Mowafa Househ. 2019. An overview of the features of chatbots in mental health: A scoping review. *International journal of medical informatics* 132 (2019), 103978.

[2] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774* (2023).

[3] Shaukat Ali, Shah Khusro, Akif Khan, Inayat Khan, and Salman Faiz Solheria. 2019. An insight of smartphone-based lifelogging research: Issues, challenges, and research opportunities. *Proceedings of the Pakistan Academy of Sciences: A Physical and Computational Sciences* 56, 3 (2019), 1–16.

[4] Alzheimer's Association. 2024. 2024 Alzheimer's Disease Facts and Figures. *Alzheimer's & Dementia* 20, 5 (2024), 3708–3821. doi:10.1002/alz.13809 Epub 2024-04-30. PMID: 38689398; PMCID: PMC11095490.

[5] Dang Anh-Hoang, Vu Tran, and Le-Minh Nguyen. 2025. Survey and analysis of hallucinations in large language models: attribution to prompting strategies or model behavior. *Frontiers in Artificial Intelligence* 8 (2025), 1622292.

[6] Elham Asgari, Nina Montaña-Brown, Magda Dubois, Saleh Khalil, Jasmine Balloch, Joshua Au Yeung, and Dominic Pimenta. 2025. A framework to assess clinical safety and hallucination rates of LLMs for medical text summarisation. *npj Digital Medicine* 8, 1 (2025), 274.

[7] Luke Balcombe. 2023. AI chatbots in digital mental health. 10, 4 (2023), 82.

[8] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems* 33 (2020), 1877–1901.

[9] Mary Butler, Ellen McCreedy, Victoria A Nelson, Priyanka Desai, Edward Ratner, Howard A Fink, Laura S Hemmy, J Riley McCarten, Terry R Barclay, Michelle Brasure, et al. 2018. Does cognitive training prevent cognitive decline? A systematic review. *Annals of internal medicine* 168, 1 (2018), 63–68.

[10] Aaron TC Chan, Roy TF Ip, Joshua YS Tran, Joyce YC Chan, and Kelvin KF Tsoi. 2024. Computerized cognitive training for memory functions in mild cognitive impairment or dementia: a systematic review and meta-analysis. *NPJ Digital Medicine* 7, 1 (2024), 1.

[11] Paramveer S Dhillon, Somayeh Molaei, Jiaqi Li, Maximilian Golub, Shaochun Zheng, and Lionel Peter Robert. 2024. Shaping human-AI collaboration: Varied scaffolding levels in co-writing with language models. In *Proceedings of the 2024 CHI conference on human factors in computing systems*. 1–18.

[12] Leah Findlater, Jon E Froehlich, Kays Fattal, Jacob O Wobbrock, and Tanya Dastyar. 2013. Age-related differences in performance with touchscreens compared to traditional mouse input. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 343–346.

[13] Alexandra M Kueider, Jeanine M Parisi, Alden L Gross, and George W Rebok. 2012. Computerized cognitive training with older adults: a systematic review. *PloS one* 7, 7 (2012), e40588.

[14] Peter Lee, Sebastien Bubeck, and Joseph Petro. 2023. Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. *New England Journal of Medicine* 388, 13 (2023), 1233–1239.

[15] Zhen Li, Hao He, Yiqi Chen, and Qing Guan. 2024. Effects of engagement, persistence and adherence on cognitive training outcomes in older adults with and without cognitive impairment: a systematic review and meta-analysis of randomised controlled trials. *Age and Ageing* 53, 1 (2024), afad247.

[16] Na Liu, Jiamin Yin, Sharon Swee-Lin Tan, Kee Yuan Ngiam, and Hock Hai Teo. 2021. Mobile health applications for older adults: a systematic review of interface and persuasive feature design. *Journal of the American Medical Informatics Association* 28, 11 (2021), 2483–2501.

[17] Our World in Data. 2024. Population aged 0–4 (high projection). https://ourworldindata.org/ Data originally from United Nations, World Population Prospects 2024. Accessed 2 March 2026..

[18] Charan Ranganath, Kristin E Flegal, and Laura L Kelly. 2011. Can cognitive training improve episodic memory? *Neuron* 72, 5 (2011), 688–691.

[19] George W Rebok, Karlene Ball, Lin T Guey, Richard N Jones, Hae-Young Kim, Jonathan W King, Michael Marsiske, John N Morris, Sharon L Tennstedt, Frederick W Unverzagt, et al. 2014. Ten-year effects of the advanced cognitive training for independent and vital elderly cognitive training trial on cognition and everyday functioning in older adults. *Journal of the American Geriatrics Society* 62, 1 (2014), 16–24.

[20] Sarah R Rudebeck, Daniel Bor, Angharad Ormond, Jill X O'Reilly, and Andy CH Lee. 2012. A potential spatial working memory training task to improve both episodic memory and fluid intelligence. *PloS one* 7, 11 (2012), e50431.

[21] Abigail J Sellen, Andrew Fogg, Mike Aitken, Steve Hodges, Carsten Rother, and Ken Wood. 2007. Do life-logging technologies support memory for the past? An experimental study using SenseCam. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 81–90.

[22] Cynthia S Symons and Blair T Johnson. 1997. The self-reference effect in memory: a meta-analysis. *Psychological bulletin* 121, 3 (1997), 371.

[23] Julie F Vermeir, Melanie J White, Daniel Johnson, Geert Crombez, and Dimitri ML Van Ryckeghem. 2020. The effects of gamification on computerized cognitive training: systematic review and meta-analysis. *JMIR serious games* 8, 3 (2020), e18644.

[24] Deborah Vollmer Dahlke and Marcia G Ory. 2017. Emerging opportunities and challenges in optimal aging with virtual personal assistants. *Public Policy & Aging Report* 27, 2 (2017), 68–73.

[25] Wiktoria Wilkowska, Julia Offermann-van Heek, Liane Colonna, and Martina Ziefle. 2020. Two faces of privacy: Legal and human-centered perspectives of lifelogging applications in home environments. In *International Conference on Human-Computer Interaction*. Springer, 545–564.

[26] Sherry L Willis, Sharon L Tennstedt, Michael Marsiske, Karlene Ball, Jeffrey Elias, Kathy Mann Koepke, John N Morris, George W Rebok, Frederick W Unverzagt, Anne M Stoddard, et al. 2006. Long-term effects of cognitive training on everyday functional outcomes in older adults. *Jama* 296, 23 (2006), 2805–2814.

[27] David Wood, Jerome S Bruner, and Gail Ross. 1976. The role of tutoring in problem solving. *Journal of child psychology and psychiatry* 17, 2 (1976), 89–100.

[28] Jie Wu, Chaofan Geng, Liyang Liu, and Yi Tang. 2025. Advancement of disease-modifying therapy of Alzheimer's disease: from the perspective of new revised criteria for diagnosis and staging of Alzheimer's disease. *Medicine Plus* 2, 4 (2025), 100112. doi:10.1016/j.medp.2025.100112

[29] Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik R Narasimhan, and Yuan Cao. 2022. React: Synergizing reasoning and acting in language models. In *The eleventh international conference on learning representations*.

[30] J Diego Zamfirescu-Pereira, Richmond Y Wong, Bjoern Hartmann, and Qian Yang. 2023. Why Johnny can't prompt: how non-AI experts try (and fail) to design LLM prompts. In *Proceedings of the 2023 CHI conference on human factors in computing systems*. 1–21.

[31] Wan Zhang and Jing Zhang. 2025. Hallucination mitigation for retrieval-augmented large language models: a review. *Mathematics* 13, 5 (2025), 856.

[32] Yue Zhang, Yafu Li, Leyang Cui, Deng Cai, Lemao Liu, Tingchen Fu, Xinting Huang, Enbo Zhao, Yu Zhang, Yulong Chen, et al. 2025. Siren's Song in the AI Ocean: A Survey on Hallucination in Large Language Models. *Computational Linguistics* (2025), 1–46.

[33] Kathrin Zimmermann, Claudia C Von Bastian, Christina Röcke, Mike Martin, and Anne Eschen. 2016. Transfer after process-based object-location memory training in healthy older adults. *Psychology and Aging* 31, 7 (2016), 798.