"Why did you say that?": Understanding Explainability in Conversational AI systems for Older Adults with Mild Cognitive Impairment (MCI)

Niharika Mathur[1], Tamara Zubatiy[1], Agata Rozga[1], Elizabeth Mynatt[2]

[1] Georgia Tech
[2] Northeastern University

Abstract. As Conversational AI systems evolve, their user base widens to encompass individuals with varying cognitive abilities, including older adults facing cognitive challenges like Mild Cognitive Impairment (MCI). Current systems, like smart speakers, struggle to provide effective explanations for their decisions or responses. This paper argues that the expectations and requirements for AI explanations for older adults with MCI differ significantly from conventional Explainable AI (XAI) research goals. Drawing from our ongoing research involving older adults with MCI and their interactions with the Google Home Hub, we highlight breakdowns in conversational flow when older adults seek explanations. Based on our experience, we conclude with recommendations for HCI researchers to adopt a more human-centered approach as we move towards developing the next generation of AI systems.

Keywords: Conversational AI · Explainable AI · Mild Cognitive Impairment · Older Adults.

1 Introduction

In this research, we discuss Conversational AI systems and their explanations in the context of supporting older adults with Mild Cognitive Impairment (MCI), an early stage of cognitive decline. These AI systems use Natural Language Processing and speech recognition to engage in real-time interactions. Notably, commercially available smart speakers, such as Google Home[3] offer natural language support to older adults by providing auditory stimuli, thus minimizing physical device engagement [6]. Older Adults with MCI have also reported feeling "empowered" through their use [8]. However, our focus in this paper is on their ability to handle more complex, personalized and interpersonal conversations that go beyond basic tasks like information retrieval and entertainment. Complex tasks, such as calendaring and information retrieval, often lead to conversational breakdowns as AI systems struggle to remember and present information effectively. Problems arise when older adults dealing with the onset of

[3] Retrieved September 24, 2023, from https://home.google.com/welcome/

changes in their cognitive abilities desire more information or explanations from the AI. This increased need for cognitive support as a result of MCI presents a challenge for the current capabilities of conversational AI systems.

We draw insights from ongoing work with this demographic, analyzing interaction logs and qualitative interviews. We argue that Explainable AI (XAI), which focuses on explaining AI decisions, holds promise in addressing these breakdowns. Our paper also critiques current approaches to designing and evaluating explanations in XAI and recommends a more inclusive approach to XAI development.

2 Background

2.1 Conversational Breakdowns in AI systems

While Conversational AI systems show potential in assisting in aging in place, there has also been concurrent research that highlights the breakdowns in the interactions that users have with AI systems, thus bringing into question the truly conversational nature of such technologies. In [1], authors Clark et al describe the current perception of conversational agents as merely task-focused entities, while highlighting the desire to incorporate the "dynamics of bond and trust" in order to make truly conversational agents that have the potential of extending their role from a functional to a social one. As we have observed in our work with older adults so far, current Conversational AI systems fall short of this expectation as interactions with them often break down or are abandoned when prompted for information besides the one that they have been programmed to store and retrieve.

2.2 Role of Explainable AI

Explainability in the context of AI systems is the ability of an AI system to provide reasoning and explanations for its decisions. A common example is when one receives recommendations with the option to know why those recommendations were made. This helps the user in understanding the context behind the recommendations and has shown to have an impact on their overall perception of the system [7]. However, most approaches to explainable AI today tend to be algorithm-centered and focused on models generated by the AI systems. The existence, type and nature of explanations provided to all users is objectively decided by the AI, with little to no understanding of an individual users' cognitive and social requirement for explanations. This highlights the need for a human-centered approach to explainability. In [2], Ehsan et al advocate for the need for social transparency in explanations by acknowledging the socially situated nature of both the AI and the people interacting with it. Through our work, we also advocate for further extending this social perspective of the user to include an examination of the cognitive perspective of users and its impact on decision making. Two users can be socially situated in the same context, however, their cognitive contexts could be vastly different. The envisioned contribution of our work comes from the need to understand explanations in AI systems in more depth in the context of older adults with MCI and their carepartners who may have different perceptions of the system and have different explainability requirements given the difference in their individual cognitive abilities.

3 Previous Work: Analyzing the use of Commercial Conversational AI systems

The context of our argument in this paper is derived from our work described in detail in [4]. To briefly summarize, our aim is to study the long-term usage of a commercial conversational AI system, the Google Home Hub (GHH), for medication management. Our research team designed a medication assistant through exploratory user research with "dyads" (i.e., pairs) of older adults with MCI and their caregivers, who are often their spouses, but also adult children in some cases. Our team recruited the dyads for the study through a cognitive empowerment program at a local healthcare facility. The purpose of the medication assistant is to check-in with the participant at a pre-defined time for their medications and follow a conversational trajectory on the basis of its current technical abilities. As detailed in [4], the team deployed the medication assistant in the GHH of 7 dyads with an average age of 74.5 for patients with MCI and 68.5 for their carepartners for a period of 20 weeks. Throughout this 20 week study period, we collected interaction logs, and conducted qualitative interviews with the dyads at two different points during the study duration. The interaction logs list and categorize the verbal interactions between a user and the GHH in a spreadsheet (each cell is a text interaction). A quantitative analysis of the interaction logs collected over 20 weeks revealed an engagement rate of 67% for all participating dyads. We define the engagement rate as the ratio of the interactions (the "check-in") that the AI initiated at the pre-defined medication time versus the interactions that the user actually responded to (or engaged in a conversation with the AI). We also calculated a weekly engagement rate for 20 weeks for all the participating dyads which showed a steady increase in engagement with the AI as the study progressed, hinting at an overall acceptance of the system. The findings from the qualitative interviews also point towards the fact that in its present form, the system provided feelings of confidence and support to the participants, specially the caregivers, who called it an *"alternate way of monitoring their partner's medication"* (Caregiver quote).

As a result of the work described in [4], we have been positively encouraged by the results hinting at an overall acceptance of the system. Our published analysis so far focuses primarily on interaction design, system usage and engagement. However, inspired from this work and continuing through recent further analysis of interaction logs and additional interviews, not previously published, we have also begun to observe an emerging trend that hints at the expectations and requirements that older adults have from the AI system that currently go unfulfilled and lead to conversational breakdowns. We discuss some of these emerging conversational breakdowns in the next section.

4 Understanding Explainability Expectations for Conversational AI

Some of the observed conversational breakdowns that hint at specific explainability expectations from Conversational AI systems are:

1. Current Conversational AI design is constrained in what it can and cannot do and works primarily on templated interactions. In the context of our work, when prompted by a user with questions such as *"Are you sure I have not taken the medication? I think I have"* (patient quote), the system has no way of presenting an explanation for its decision to clarify user skepticism, instead answering with the same template response with no added information about an individual patient's medication status. This highlights the conversational limitation for the AI in its explanations to a user.
2. In most cases, the caregivers or the patients want to know more about their medication history for personal tracking or sharing medication records with clinicians. One caregiver asked the system *"Can you tell me if he took the medication this morning or no?"*, but was met with no response as the system has no internal knowledge about previous interactions, hinting at its lack of explainability and information storing capabilities.
3. The requirement for more information and explanation from the system varies across users. Some never asked a follow-up question and assumed the AI must be right at all times, while others wanted to know more about their medication data and for the AI to provide more explanation for its responses. For the latter case, a patient describes, *"Can it tell me for sure that I have not taken the medication because I think I have and there is no pill in the pillbox too"*. The system's non-response to user questions like this highlights the expectation for the AI to calibrate responses to a user's cognitive model in order to effectively build trust in it.

4.1 Conversational breakdowns through the lens of Explainable AI

Through analysis of our results so far, we argue that the emergence of conversational breakdowns and user frustration is closely tied to a lack of explainability in Conversational AI systems. Our argument in this paper is rooted in the ways in which current conversational AI systems are inept at working with complex interactions and to contribute to the understanding of how they can be better designed in the future. There is currently a gap in the way that explanations are understood in the context of non-traditional users interacting with conversational AI systems. We further argue that not every explanation offered is a good one, the central concern is whether it serves the purpose of explainability for a specific user in the context of their social and cognitive abilities. We highlight that explainability requirements are different for older adults and we need to take a significantly different, human-centered approach to understanding explainability than what most XAI studies presently do.

Designing explanations for AI systems requires understanding aspects that relate to define what an effective explanation is, when it is delivered and how it can impact a user's mental model [3]. Currently, explanations in XAI studies are largely constructed on machine-generated inputs. The generated explanation in most XAI studies has not been trained on the system's internal knowledge of a user interacting with it. The nature of explanations provided to all users is objectively decided by the AI, with little to no understanding of an individual user's cognitive model. This mismatch in a user's expectation from an explanation and the actual explanation can influence trust in the system [7]. User frustrations resulting from ambiguity in explanations can lead to reduced feelings of autonomy and independence. These are two central concerns for older adults as they age with cognitive impairments [5]. We also highlight that in XAI studies, the generated explanations are largely tested with low-risk tasks that are not representative of real-world scenarios. The role of risk associated with a task has also not been explored so far. Additionally, these evaluations are often performed by standalone non-experts who test the system in isolation and not in collaborative contexts.

4.2 Recommendations for Explainable AI studies

Concluding our argument so far, we offer three broad recommendations for HCI researchers working towards understanding explanations in the next generation of AI systems.

R1: Explanations in XAI studies need to be generated keeping in mind a user's individual cognitive and social context. Conducting exploratory user research with actual users of the system can be helpful in shifting the algorithmic-centeredness of explanations to the mental model of the user. Generation of different levels of explanations that vary by the level of information contained in them or by the type of information (visual, non-visual) is also important to calibrate explanations with real-world scenarios.

R2: Explanations in XAI studies need to be evaluated in the context of use-inspired scenarios that are representative of real-world situations. A diversity in tasks that includes understanding differing perceptions of risk involved in them is crucial in grounding explainability in more authentic, collaborative and networked contexts rather than in isolation. Here, we highlight that explanations for high-risk tasks such as medication reminders may require a different evaluatory plan than a low-risk such as movie recommendation.

R3: Finally, we offer the recommendation that XAI studies should engage more representative users for performing experiments rather than users that can be conveniently recruited. In this context, evaluations with experts in their life experience of living with MCI has the potential to lead to the generation of more actionable insights for the design of XAI systems in the future.

5 Conclusion and Future Work

In this paper, we have presented an overview of our ongoing work with older adults with MCI and their use of Conversational AI systems. As older adults continue to age, there is a pressing need to develop systems that can engage more robustly in order to truly provide support for aging in place. We argue that this robustness is closely tied to building effective explanations into AI systems that can enhance conversational interactions. We further argue that explainability for older adults requires a human-centered approach that keeps them at the center of the explanation. We conclude this work in progress by offering recommendations to generate cognitively situated explanations, to test them with diverse, risk-aware tasks and with users that truly represent the use context representative of the scenario in question. Our future work focuses on evaluating different levels and types of explanations with older adults and representative users. Through the use user-centered qualitative research, we plan to conduct an in depth analysis of what AI explainability means for diverse users.

References

1. Clark, L., Pantidi, N., Cooney, O., Doyle, P., Garaialde, D., Edwards, J., Spillane, B., Gilmartin, E., Murad, C., Munteanu, C., et al.: What makes a good conversation? challenges in designing truly conversational agents. In: Proceedings of the 2019 CHI conference on human factors in computing systems. pp. 1–12 (2019)
2. Ehsan, U., Liao, Q.V., Muller, M., Riedl, M.O., Weisz, J.D.: Expanding explainability: Towards social transparency in ai systems. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. pp. 1–19 (2021)
3. Liao, Q.V., Gruen, D., Miller, S.: Questioning the ai: informing design practices for explainable ai user experiences. In: Proceedings of the 2020 CHI conference on human factors in computing systems. pp. 1–15 (2020)
4. Mathur, N., Dhodapkar, K., Zubatiy, T., Li, J., Jones, B., Mynatt, E.: A collaborative approach to support medication management in older adults with mild cognitive impairment using conversational assistants (cas). In: Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility. pp. 1–14 (2022)
5. Rogers, W.A., Mitzner, T.L.: Envisioning the future for older adults: Autonomy, health, well-being, and social connectedness with technology support. Futures **87**, 133–139 (2017)
6. Sengupta, K., Sarcar, S., Pradhan, A., McNaney, R., Sayago, S., Chattopadhyay, D., Joshi, A.: Challenges and opportunities of leveraging intelligent conversational assistant to improve the well-being of older adults. In: Extended abstracts of the 2020 CHI conference on human factors in computing systems. pp. 1–4 (2020)
7. Shin, D.: The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable ai. International Journal of Human-Computer Studies **146**, 102551 (2021)
8. Zubatiy, T., Vickers, K.L., Mathur, N., Mynatt, E.D.: Empowering dyads of older adults with mild cognitive impairment and their care partners using conversational agents. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. pp. 1–15 (2021)