Exploring Photo-Based Dialogue Between Elderly Individuals and Generative AI Agents

Kousuke Shimizu[1], Banba Ami[2], Choi Dongeun[3], Miyuki Iwamoto[4],
Nahoko Kusaka[5], Panote Siriaraya[6], Noriaki Kuwahara[7]*
Graduate School of Science and Technology, Kyoto Institute of Technology, Kyoto, Japan[1, 2, 6, 7]
Department of Informatics, The University of Fukuchiyama, Kyoto, Japan[3]
Department of Social System Studies, Doshisha Women's College of Liberal Arts, Kyoto, Japan[4, 5]

***Abstract*—Japan's rapid transition into a super-aged society, with 29% of its population aged 65 and over, underscores the urgent need for innovative elderly care solutions. This study explores the use of generative AI to facilitate meaningful interactions between elderly individuals and AI conversational agents using photos. Utilizing Microsoft Azure's AI services, including Computer Vision and Speech, the AI agent analyzes photos to generate engaging conversation prompts, leveraging GPT-3.5-turbo for natural language processing. Preliminary experiments with healthy elderly participants provided insights to refine the AI agent's conversational skills, focusing on timing, speech speed, and emotional engagement. The findings indicate that elderly users respond positively to AI agents that exhibit human-like conversational behaviors, such as attentiveness and expressive communication. By addressing functional and emotional needs, the AI agent aims to enhance the quality of life for the elderly, offering scalable solutions to the challenges of an aging society. Future work will focus on further improving the AI agent's capabilities and assessing its impact on the mental health and social engagement of elderly users.**

Keywords—Generative AI; elderly care; conversational agents; photo-based interaction

I. Introduction

Dementia care demands specialized knowledge and experience, yet the increasing shortage of caregivers makes it difficult to provide the necessary personalized attention. This gap in care contributes to feelings of isolation and anxiety among the elderly, exacerbating behavioral and psychological symptoms associated with dementia (BPSD), such as agitation, depression, and social withdrawal [1]. Individuals are exploring innovative solutions leveraging information and communication technology (ICT) to address these challenges. While previous studies have explored various technological interventions, there remains a significant gap in personalized, scalable solutions that can effectively reduce the burden on caregivers and enhance the quality of life for elderly individuals [2]. This study focuses on the use of AI conversational agents to support the elderly through meaningful interactions. By employing advanced AI technologies, particularly those capable of natural language processing and responsive interaction, these agents can potentially reduce the burden on caregivers and enhance the quality of life for elderly individuals.

This research aims to develop an AI agent using Microsoft's Azure services, such as Computer Vision and Speech, and the GPT-3.5-turbo language model. We designed the agent to engage elderly users in conversations based on photo prompts, providing companionship and cognitive stimulation. Previous studies have shown the effectiveness of conversational AI and robots in elderly care and dementia support [3][4]. However, these studies often lack personalization and real-time adaptability, which are crucial for effectively supporting elderly individuals with varying needs.

The initial phase of the study involves healthy elderly participants to develop an understanding of general communication patterns and preferences. The insights gained will guide the development of the AI agent, ensuring it can effectively mimic human-like conversational behaviors and meet the social and emotional needs of its users.

Ultimately, the goal is to demonstrate that AI agents can be a viable solution for enhancing elderly care, offering scalable and effective support to address the challenges posed by Japan's rapidly aging society. Future research will focus on refining the AI agent's capabilities and evaluating its impact on the mental health and social engagement of elderly users. The development of AI agents like ours is also supported by recent research which highlights their potential in improving the mental health and well-being of elderly individuals [5] [6].

II. Related Work

The field of human-robot interaction has seen significant advancements, particularly in enhancing communication with elderly users. Yoshida et al. studied the impact of robots' behaviors during conversational pauses, finding that natural gestures by robots can make these pauses feel shorter and improve the overall flow of conversation [7]. The study suggests that human-like behaviors in robots can enhance the user experience. Similarly, conversational AI has been successfully employed to alleviate loneliness and promote social interaction among the elderly [8][9].

Heerink et al. explored how the social capabilities of robots affect their acceptance among elderly users [10]. They identified key social behaviors such as attentiveness, positive communication, personal references, expressiveness, and the ability to admit mistakes. Their findings indicate that robots with these capabilities are more likely to be perceived as comfortable and engaging communication partners. In line with these findings, Mendes et al. demonstrated that emotionally intelligent avatars could enhance elderly care in ambient assisted living environments [11].

In the context of dementia care, N. Saito et al. present the development of a multimodal conversational agent system designed to interact with elderly patients with dementia [12]. The primary aim is to improve the system's capability to recognize when a subject has the right to speak based on cues from their spontaneous speech and other modalities such as gaze and head motion. This mechanism is crucial for facilitating smoother and more intuitive interactions. The paper outlines a turn-taking strategy that utilizes these cues to interpret pauses in speech better, which are frequent in dialogues with dementia patients. This highlights the importance of designing AI agents that can adapt to the specific communication needs of dementia patients.

These studies collectively emphasize the potential of AI and robotic technologies to improve social interaction and support for elderly individuals, particularly those with cognitive impairments. They provide a foundation for further research into the development of AI agents that can effectively engage elderly users through natural and empathetic communication. Additionally, recent research has focused on using conversational agents and robots to support the well-being of elderly individuals, demonstrating positive outcomes in emotional and cognitive engagement [13] [14]. However, these studies often lack a focus on personalized, photo-based interactions which can provide more relevant and engaging experiences for the elderly.

It has been clearly demonstrated that dialogues involving content such as photographs can reduce the conversational burden on young caregivers when interacting with elderly patients with dementia [15-17]. Additionally, using photographs from memories in conversations with elderly dementia patients can also have a reminiscence therapy effect [18]. Based on these studies, we have decided to research and develop a conversational agent that can understand photographs brought by the elderly and engage in meaningful discussions about them with the elderly. Furthermore, the use of AI-driven interactive multimodal photo albums has shown promise in enhancing personalized reminiscence therapy among older adults [19] [20].

Our research aims to bridge the gap by developing a photo-based conversational AI agent that not only facilitates meaningful interactions but also provides emotional support and cognitive stimulation, thereby enhancing the overall quality of life for elderly users. The novelty of our approach lies in integrating advanced AI technologies with photo-based dialogue, leveraging recent advancements in AI and conversational agents to create a more engaging and supportive environment for elderly care.

III. Conversational Agent Using Generative AI

This chapter outlines the conversational agent’s technical framework and system architecture developed using generative AI technologies. The agent leverages several advanced AI services that Microsoft Azure provides to facilitate natural and meaningful interactions with elderly users.

A. Technologies Used

1) Microsoft azure: Azure provides a robust cloud computing platform that supports various AI services essential for developing and deploying the conversational agent.

2) Computer Vision (CV): The Computer Vision service, provided by Microsoft Azure, analyzes photos to generate captions and identify objects within the images. This information is used to create relevant and engaging conversation prompts for the AI agent.

3) Speech service: Azure's Speech service enables the agent to convert text to natural-sounding speech and vice versa. This functionality is crucial for real-time, spoken interactions with users.

4) OpenAI service: The conversational agent utilizes the GPT-3.5-turbo language model from OpenAI, integrated through Microsoft Azure, to generate human-like responses. This model is capable of understanding and producing text that is contextually relevant and coherent.

B. System Architecture and Implementation

1) Photo analysis: When a user provides a photo, the Computer Vision service analyzes the image and generates descriptive captions. These captions highlight key elements and contexts within the photo.

2) Question generation: The generated captions are then fed into the GPT-3.5-turbo model, which creates relevant questions and conversational prompts based on the photo's content. This step ensures that the conversation remains engaging and contextually appropriate.

3) Speech interaction: The conversational prompts are converted to speech using the Azure Speech service, allowing the AI agent to communicate verbally with the user.

4) Real-time response: The user’s spoken responses are transcribed into text, which the GPT-3.5-turbo model processes to generate appropriate replies. These replies are then converted back into speech, creating a seamless conversational experience.

5) Prompt setting: The prompts used by the AI agent are carefully designed to be clear and contextually relevant. This involves setting specific parameters in the GPT-3.5-turbo model to ensure the generated questions and responses are engaging and appropriate for elderly users.

6) Speech synthesis customization: The Speech service's output is customized to produce natural and empathetic voice tones. This customization includes adjusting prosody and phoneme settings to make the AI agent’s speech more pleasant and understandable for elderly users.

7) Iterative improvement: The AI agent’s responses are tailored based on user feedback to optimize conversation timing, speech speed, and emotional engagement. This iterative improvement process ensures that the AI agent remains responsive and supportive during interactions.

This architecture and implementation strategy enables the development of an AI conversational agent that not only supports the cognitive and emotional needs of elderly users but also provides scalable solutions for enhancing elderly care in a rapidly aging society. Fig. 1 illustrates the conversational agent's system configuration.

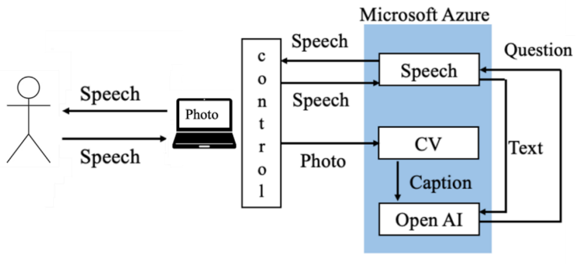

Fig. 1. Illustrative example of the conversational agent's system configuration.

IV. Methods

This chapter describes the study's methodology, including the preliminary studies and the main experiment conducted with elderly participants from Kyoto Institute of Technology (KIT).

A. *Preliminary Study*

The preliminary study involved two experiments to gather insights for refining the AI conversational agent. The first experiment was conducted in collaboration with Doshisha Women's College of Liberal Arts, and the second involved student participants.

1) Experiment I: Collaboration with Doshisha Women's College

a) Objective: To observe interactions between AI agents and elderly users in an uncontrolled, real-world environment.

b) Reason for selection: Doshisha Women's College was chosen due to its active engagement in community service and research on elderly care, providing a diverse and relevant participant pool.

c) Method: Participants interacted with the AI agent in their natural settings, and their conversations were recorded and analyzed.

d) Findings: The study identified key conversational patterns and user preferences, such as the importance of conversation timing, speech speed, and emotional engagement.

2) Experiment II: Student Experiment

a) Objective: To further refine the AI agent's responses based on feedback from younger participants.

b) Method: Student interactions with the AI agent were recorded and analyzed like the first experiement.

c) Findings: The study provided additional insights into the agent's performance, highlighting areas for improvement in natural language processing and responsiveness.

3) Improvements made to the AI agent based on preliminary study

a) Utterance length: Shortened the length of each utterance from 100 characters to 50 characters to make the conversation more concise and easier to follow.

b) Number of questions per utterance: Limited the agent to ask only one question per utterance to simplify interactions and avoid overwhelming the user.

c) Response delay: Adjusted the waiting time to up to 10 seconds at the start of a conversation and five seconds during conversation stalls to allow users more time to respond.

d) Speech speed: The speech speed of the AI agent was fine-tuned to match the preferred pace of elderly users, making it easier for them to engage in the conversation. Observing that some elderly participants were urging the agent to respond more quickly, the speed of the AI agent's speech was slightly increased using the prosody rate.

e) Name recognition: The elderly participants were asked for their names, and the agent used their names during the conversation to foster a sense of attachment to the agent. Improved the agent's name recognition by asking participants to provide only their given names instead of full names. The agent included a self-introduction ("Nice to meet you. I am Aiko. Please tell me your given name") to create a natural flow for name exchange.

Fig. 2. An elderly person interacting with the AI agent.

Fig. 2 is a photo showing an elderly person interacting with the AI agent.

B. *Main Experiment with Elderly Participants*

The main experiment aimed to evaluate the effectiveness of the improved AI conversational agent in real-world settings with elderly participants.

1) Experiment overview: The main experiment aimed to explore methods to make interactions between the AI agent and elderly users more natural and stress-free and to verify how closely the agent could emulate human conversation. With the cooperation of the Silver Human Resources Center, 12 elderly participants aged 65 and over took part in the experiment. The sample size was determined based on preliminary studies and practical constraints. Additionally, one student participated to act as a comparison for the conversations between the agent and humans. In the experiment, elderly participants first interacted with Agent 1 (before Improvements) for about two minutes, discussing a photo. After this conversation, they answered a questionnaire about their experience with Agent 1. They then had a similar two-minute conversation with Agent 2 (after Improvements), followed by another questionnaire. Finally, they conversed with the student about the same photo for comparison.

2) Evaluation method: Participants answered a questionnaire after each interaction with the agents, evaluating eight aspects on a five-point scale:

a) Overall satisfaction with the conversation (OS)

b) Whether the conversation was well-established (WE)

c) Smoothness of responses (SMR)

d) Naturalness of the interaction (NAI)

e) Enjoyment of talking (EJT)

f) Presence of emotions (PE)

g) Understanding of speech (US)

h) Willingness to talk again (WT)

Questionnaire items were designed with reference to research by Yoshida et al. [7]. Feedback from experts in elderly care and AI interaction was incorporated to refine the questions. Participants rated these aspects from 1 (strongly disagree) to 5 (strongly agree). Additionally, they provided free-text comments about their experience with each agent. This comprehensive evaluation helped assess the effectiveness and user satisfaction of the conversational agents.

We recorded and transcribed the conversations to count the number of utterances made by participants. Utterance units were counted up to a period mark, and continuous phrases like "yes, that’s right" were counted as one unit. Interjections like "uh" were not included in the count. Additionally, the frequency of positive and negative body language and the number of overlaps in conversation with the agent were measured. Positive body language included smiles and nods in response to the agent's speech, while negative body language included frowning, tilting the head, covering the mouth, and resting elbows on the table.

V. Results

A. Results of Post-Experiment Questionnaire

Fig. 3 shows the results of the Post-experiment Questionnaire. To compare the mean scores of each metric among the three groups (Agent1, Agent2, and Human), an ANOVA was conducted, followed by post hoc tests to examine whether there were significant differences in the mean scores. Fig. 3 indicates significant differences, with ** denoting p<0.01 and * denoting p<0.05. The following summarizes the results of comparing various conversational metrics between the pre-improvement agent (Agent1), the post-improvement agent (Agent2), and humans.

1) Overall Satisfaction (OS): While the differences in overall satisfaction were not statistically significant among the groups, the trend indicates a higher satisfaction rate for the Human group at 4.67, compared to 4.00 for Agent1 and 3.83 for Agent2. This suggests that while the agents are capable of achieving a satisfactory level of interaction, they still fall short of human standards.

2) Well-Established Conversation (WE): Significant differences were noted in terms of establishing a well-rounded conversation. The Human group scored the highest (4.83), significantly outperforming Agent1 (3.58). Although Agent2 (4.08) showed an improvement over Agent1, it still did not match the human interaction quality, indicating room for enhancement in AI conversational frameworks.

3) Smoothness of Responses (SMR): Similarly, the Human group led with a score of 4.75, reflecting smoother and more coherent interactions than those managed by Agent1 (3.67) and Agent2 (3.92). This dimension particularly highlights the challenges AI agents face in replicating the fluid and adaptive nature of human responses.

4) Naturalness of Interaction (NAI): The Human group achieved the highest naturalness score (4.83), with significant differences observed when compared to both Agent1 (3.83) and Agent2 (3.58). The lower score for Agent2 suggests that despite efforts to enhance conversational mechanics, achieving a natural flow in AI-driven interactions remains a critical challenge.

5) Enjoyment of Talking (EJT): Enjoyment levels were significantly higher in human interactions (Human group, 4.75) compared to the uniform scores of 3.75 for both agents. This indicates that while AI agents can facilitate functional conversations, they are less successful at engaging users on a more personal and enjoyable level.

6) Presence of Emotions (PE): Reflecting on emotional engagement, the Human group scored 4.67, significantly higher than both Agent1 (3.42) and Agent2 (3.50). This underscores the difficulty AI agents encountered in effectively mimicking human emotional cues, which are essential for more empathetic and engaging interactions.

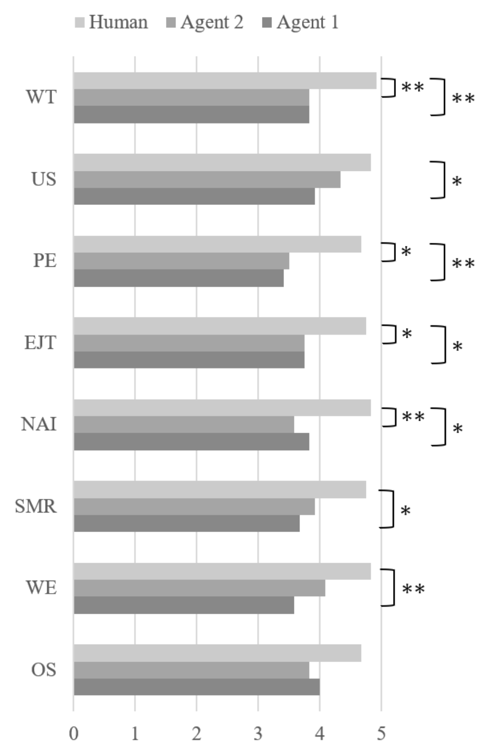

Fig. 3. The results of post-experiment questionnaire.

7) Understanding of Speech (US): The ability to comprehend speech saw Agent2 (4.33) closing the gap slightly with the Human group (4.83), compared to Agent1 (3.92). This suggests that the modifications made to Agent2, perhaps in processing or delivering speech, had a positive impact, enhancing understanding among users.

8) Willingness to Talk Again (WT): Reflecting participants' readiness to re-engage, the Human group scored highest at 4.92, significantly exceeding the scores of both Agent1 and Agent2, each at 3.83. This result highlights a critical aspect of user experience, where, despite technological advancements, human interactions remain more appealing and rewarding.

B. Results of Video Analysis During Conversations

1) Analysis of utterance counts in conversations with AI agents: We quantified the utterances during the conversations and computed the average counts for each agent. We designed the enhancements to Agent2 to improve user interaction, potentially influencing the frequency and quality of user responses. We utilized a paired t-test to assess statistically significant differences in utterance counts between the conversations with the two agents.

The findings showed that participants had more frequent utterances in conversations with Agent2, averaging 10.17 (SD=3.29), compared to 7.42 (SD=2.31) with Agent1. The statistical analysis indicated a significant difference (p=0.016), confirming that the enhancements in Agent2 facilitated more active engagement from the participants.

2) Analysis of speech overlaps in conversations with AI agents: We extracted data regarding the number of times speech overlaps occurred for each participant with both agents. The study included interactions with 12 participants, and from transcribed video recordings of these interactions, we counted the speech overlap events.

The mean number of overlaps per conversation was calculated for every interaction with each agent. For Agent1, the mean overlap count was approximately 1.75, whereas for Agent2, it was about 0.92. These results were analyzed statistically using a paired t-test, which yielded a t-value of 1.52 and a p-value of 0.157.

3) Analysis of positive and negative gestures in conversations with dialogue agents: We extracted data on the count of positive and negative gestures for each participant with both agents from the video-recorded sessions and calculated the mean counts for both types of gestures across interactions with each agent.

The mean counts for positive gestures were approximately 0.92 for Agent1 and 1.92 for Agent2. For negative gestures, the means were 1.00 for Agent1 and 0.67 for Agent2. Statistical analysis using paired t-tests showed a t-value of -2.17 with a p-value of 0.053 for positive gestures and a t-value of 1.00 with a p-value of 0.339 for negative gestures.

VI. Discussions

A. Assessing AI Agent Improvements Based on Questionnaire Results

Based on the improvements implemented in Agent2 and the results from our analysis, we can evaluate the effectiveness of these modifications relative to Agent1. The key areas of enhancement include utterance length, number of questions per utterance, response delay, speech speed, and name recognition.

1) Utterance length: While there were no significant differences in overall satisfaction, the enhancements might have contributed to a higher perceived naturalness and ease of following the conversation. However, Agent2 still did not significantly outperform Agent1 in naturalness, suggesting that while shorter utterances are beneficial, they alone may not be sufficient to dramatically enhance user experience. Zierau et. al., emphasizes the importance of semantic fluency and conversational design in voice-based interfaces. It suggests that limiting the number of conversational turns and using simpler, more familiar words enhances the user experience. Our finding on utterance length complements this by showing that shorter utterances can contribute to perceived naturalness and ease of conversation. This highlights the need for a holistic approach to conversational design, considering both utterance length and other factors such as contextual relevance and emotional engagement [23].

2) Number of questions per utterance: This change likely helped improve the structure and flow of conversations, potentially contributing to the increased scores in "Well-Established Conversation" for Agent2 compared to Agent1. Simplifying interactions in this way can be particularly effective in settings involving complex information or users who may benefit from a more straightforward communication style. This aligns with the importance of optimizing conversational turns [23], as noted in our finding on Utterance Length. By simplifying interactions and reducing cognitive load, the improved structure and flow likely enhanced user experience, especially for those needing straightforward communication.

3) Response delay: Although specific metrics related to response delay weren't directly measured, this modification likely enhanced the user's comfort and satisfaction with the pace of the conversation, as reflected Agent2's slightly better scores in understanding speech. Allowing users more time to think and respond can be crucial for maintaining a fluid and stress-free dialogue, especially with elderly users. Although specific metrics related to response delay weren't directly measured, this modification likely enhanced the user's comfort and satisfaction with the pace of the conversation, as reflected in Agent2's slightly better scores in understanding speech. Allowing users more time to think and respond can be crucial for maintaining a fluid and stress-free dialogue, especially with elderly users. Optimizing response delay, like limiting conversational turns and simplifying language, helps create a more user-friendly and engaging experience [23]. This complements our findings on Utterance Length and Number of Questions per Utterance, highlighting the importance of both content and pacing in conversational design to improve user satisfaction and understanding.

4) Speech speed: The fine-tuning of speech speed to match user preferences and the responsive adjustment based on user feedback may have contributed to Agent2's improved performance in "Understanding of Speech". These results may indicate that tailoring speech dynamics to the audience can significantly improve communication effectiveness. Our findings on speech speed align with Christenson et al.'s research, which indicates that tailored speech rates enhance user interaction with digital assistants. It suggests that adjusting speech dynamics can significantly enhance communication effectiveness and user experience [24].

5) Name recognition: Using participants' names likely contributed to higher scores in "Willingness to Talk Again" and potentially in "Enjoyment of Talking," as it personalizes the interaction. This enhancement can make conversations feel more engaging and tailored to the individual, fostering a greater sense of connection between the user and the agent. This finding is supported by the work of Alessa and Al-Khalifa, who demonstrated that personalized interactions, including the use of users' names, significantly enhance the perceived quality of conversational agents [6].

B. Assessing AI Agent Improvements Based on Video Analysis Results

This section discusses the efficacy of various improvements implemented in Agent1, evaluated through an analysis of video-recorded dialogues. The evaluation was based on quantitative and qualitative analysis of video recordings, examining changes in utterance counts, speech overlaps, and the frequency of positive and negative gestures.

1) Utterance length: The intention behind reducing utterance length was to make the dialogues more concise and manageable. This change aimed to facilitate easier comprehension and smoother turn-taking, potentially leading to more dynamic and engaging conversations. Our approach is consistent with previous research indicating that shorter, clearer and more thought-out utterances can enhance the clarity and fluidity of interactions, as observed by Pou-Prom et al. in their study on conversational robots for Alzheimer's patients [21].

2) Number of questions per utterance: Limiting the number of questions asked by the agent in each utterance sought to simplify the interactions and reduce the cognitive load placed on the users. This approach was hypothesized to decrease interruptions and speech overlaps, enhancing conversational clarity and user comfort. This aligns with findings from Stara et al., who found that reducing complexity improved user engagement and satisfaction [13].

3) Response delay: Adjusting the timing of the agent's responses was designed to give users adequate time to think and respond, which is crucial for maintaining a natural interaction rhythm. This modification was expected to reduce rushed responses and speech overlaps, contributing to a more relaxed dialogue atmosphere. The importance of conversational design is highlighted in ensuring user-friendly interactions. Our findings on response delay complement those of Zierau et. al. by emphasizing the need for adequate response timing, reducing rushed interactions, and creating a relaxed dialogue atmosphere, which aligns with optimizing semantic fluency and interaction rhythm. [23].

4) Speech speed: The optimization of speech speed to align with user preferences, mainly catering to elderly participants, was anticipated to improve engagement. Proper pacing is essential for users to process the information and participate actively in the conversation fully. This is in agreement with the study by Valtolina and Hu, which demonstrated that tailored speech speed significantly enhances the accessibility and enjoyment of interactions for elderly users [8].

5) Name recognition: Enhancing the agent's ability to recognize and use the user's name was intended to personalize the interaction, making it feel more tailored and respectful. This personal touch was expected to increase positive gestures, which indicates higher user satisfaction and comfort during the dialogue. The importance of personalized interactions, including name recognition, has been highlighted in previous studies, such as the work by Alessa and Al-Khalifa [6].

C. Summary

The reviewed document [22], focusing on dialogue systems, discusses various aspects relating to assessing AI agent improvements based on questionnaire results. While this review provides a broad review of dialogue systems, including techniques that could be applied to enhance AI agent interactions as described in your assessment points, it does not specifically address all the detailed aspects such as utterance length or response delay metrics. However, its coverage of dialogue system enhancements and capabilities could indirectly apply to analyzing AI agent performance improvements based on questionnaire results.

Also, the document's coverage of various technologies and strategies in dialogue systems provides a theoretical basis for understanding how such improvements, as observed through video analysis, can affect user interactions. The details on managing interaction dynamics, personalization, and response adaptation discussed can be considered relevant to analyzing AI agent improvements based on video analysis results. Overall, our findings indicate that personalized and context-aware enhancements in conversational AI can significantly improve user experience, aligning with broader trends in the literature on human-computer interaction [25].

VII. Limitations of this Study

A. Challenges in Achieving Natural Dialogue

Despite efforts to simulate natural dialogue, the AI agents did not fully achieve a level of interaction comparable to human conversations. The ability of AI to adequately recognize and respond to emotional cues remains limited, impacting the overall interaction quality and user experience.

B. Applicability of Statistical Analysis

Some metrics in this study did not show significant differences, possibly due to the small sample size or suboptimal statistical methods. These results make assessing the AI agent's effectiveness accurately more complicated and could skew the understanding of its impact.

C. Lack of Long-term Evaluation

The study primarily focused on short-term interactions and did not explore the long-term impacts of using AI agents on the elderly's quality of life. Long-term effects are crucial to fully understand how continuous interaction with AI influences elderly individuals' mental health and social engagement.

D. Use of Outdated AI Technology

The research utilized the GPT-3.5-turbo model and did not incorporate the latest advancements, such as the multimodal GPT-4-o. This may limit the study's relevance as newer models might offer better performance in natural language understanding, multimodal capabilities, and emotional intelligence, which are essential for enhancing interactions in eldercare.

VIII. Conclusion

Based on the findings and discussions presented in the paper, here is a conclusion in English:

This study explored innovative approaches to eldercare using dialogues facilitated by conversational AI agents through photo-based interactions. By utilizing Microsoft Azure's AI services, we developed an AI agent capable of generating conversation prompts from photos provided by elderly users. We aimed to enhance meaningful interactions and reduce their feelings of social isolation, thus improving their quality of life. Our approach addresses the gap in existing research by providing a personalized, scalable solution that leverages recent advancements in AI technology [6].

Through preliminary and main experiments, the AI agent demonstrated improvements in conversational naturalness, response smoothness, and emotional engagement. However, despite these advancements, the agent still fell short compared to human interactions, particularly in aspects like naturalness and emotional presence. This underscores the ongoing challenge of achieving human-like interaction quality in AI agents, as noted in previous studies [6].

We assessed the improvements made to Agent1 to determine their impact on enhancing the user experience during interactions. From the analysis of the questionnaire survey, the improved Agent2 have shown some effectiveness in enhancing user experience compared to Agent1, particularly in making interactions more structured, personalized, and responsive to user needs. However, the enhancements did not uniformly elevate Agent2 to the level of human interactions (Human group), indicating that while the changes are steps in the right direction, there is still significant room for improvement. Future research should continue to focus on refining these aspects to better emulate the nuanced and adaptive nature of human conversation [13].

The analysis of video-recorded dialogues suggested that these enhancements likely contributed to a more efficient and user-friendly conversational environment. Further research could explore the individual effects of each enhancement in more detail to refine the agent's capabilities and better suit user needs. Additionally, integrating multimodal cues and improving emotional recognition could further enhance the effectiveness of AI agents in elderly care [12].

Nevertheless, the interventions by AI agents showed potential benefits for the mental health and social engagement of elderly individuals, indicating that further research and refinement of the AI agent's capabilities are necessary. Future studies should focus on making the AI agent more sensitive to the emotional and social needs of elderly users and enhancing its ability to conduct dialogues that are as natural and human-like as possible. Our findings support the broader trend in AI research, which emphasizes the importance of personalized and emotionally intelligent interactions for improving user satisfaction and engagement [11].

This research underscores the potential of AI and conversational agents as effective tools for enhancing the quality of eldercare. As AI technology continues to evolve, leveraging these advancements in eldercare solutions will become increasingly important, especially in societies facing significant aging populations like Japan. The practical implications of our work suggest that AI agents can play a crucial role in providing scalable and personalized care solutions, thereby addressing the challenges posed by an aging society [5].

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number JP 24K02977.

References

[1] N. Kameoka, S. Sumitani, and T. Ohmori, “Behavioral and psychological symptoms of dementia (BPSD) and care burden: Examination in the facility staff for elderly residents”. The Journal of Medical Investigation, 67(3.4), 2020, pp.236-239. https://doi.org/10.2152/jmi.67.236

[2] Han Li, Renwen Zhang, Yi-Chieh Lee, Robert E. Kraut, David C. Mohr, “Systematic review and meta-analysis of AI-based conversational agents for promoting mental health and well-being”, npj Digital Medicine, 6:236, 2023. https://doi.org/10.1038/s41746-023-00979-5

[3] Lima, M. R., Horrocks, S., Daniels, S., Lamptey, M., Harrison, M., & Vaidyanathan, R. (2023, August). The Role of Conversational AI in Ageing and Dementia Care at Home: A Participatory Study. In 2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN) (pp. 571-578). IEEE. https://doi.org/10.1109/RO-MAN57019.2023.10309459

[4] Gilman, E. S., Kot, S., Engineer, M., & Dixon, E. (2024, March). Training Adults with Mild to Moderate Dementia in ChatGPT: Exploring Best Practices. In Companion Proceedings of the 29th International Conference on Intelligent User Interfaces (pp. 101-106). https://doi.org/10.1145/3640544.3645230

[5] Hong, Junyuan, Wenqing Zheng, Han Meng, Siqi Liang, Anqing Chen, Hiroko H. Dodge, Jiayu Zhou, and Zhangyang Wang. "A-CONECT: Designing AI-based Conversational Chatbot for Early Dementia Intervention." In ICLR 2024 Workshop on Large Language Model (LLM) Agents. https://openreview.net/forum?id=rACfuoNKBU

[6] Alessa, Abeer, and Hend Al-Khalifa. "Towards designing a ChatGPT conversational companion for elderly people." In Proceedings of the 16th international conference on Pervasive technologies related to assistive environments, pp. 667-674. 2023. https://arxiv.org/pdf/2304.09866

[7] Takumi Yoshida, Yasutake Takahashi, and Satoki Tsuichihara, “Effect of Humanoid Robot’s Response during Conversation Blank”, The 36th Fuzzy System Symposium, 2020; https://doi.org/10.14864/fss.36.0_13

[8] Valtolina, S., & Hu, L. (2021, July). Charlie: A chatbot to improve the elderly quality of life and to make them more active to fight their sense of loneliness. In Proceedings of the 14th Biannual Conference of the Italian SIGCHI Chapter (pp. 1-5). https://doi.org/10.1145/3464385.3464726

[9] Hossain, G., Pomare, Z. S., & Prybutok, G. (2024, January). ChatGPT: A Companion for Dementia Care. In 2024 IEEE International Conference on Consumer Electronics (ICCE) (pp. 1-6). IEEE. https://doi.org/10.1109/ICCE59016.2024.10444253

[10] M. HEERINK, B. KROSE, V. EVERS and B. WIELINGA, “Studying the acceptance of a robotic agent by elderly users”, International Journal of ARM, VOL. 7, NO. 3, September 2006. https://mheerink.home.xs4all.nl/pdf/IJARM.pdf

[11] Mendes, C., Pereira, R., Ribeiro, J., Rodrigues, N., & Pereira, A. (2023, July). Chatto: An emotionally intelligent avatar for elderly care in ambient assisted living. In International Symposium on Ambient Intelligence (pp. 93-102). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-43461-7_10

[12] N. Saito, S. Okada, K. Nitta, Y. Nakano, and Y. Hayashi, “Estimating user's attitude in multimodal conversational system for elderly people with dementia”, In 2015 AAAI spring symposium series; https://cdn.aaai.org/ocs/10274/10274-45306-1-PB.pdf

[13] Stara, Vera, Benjamin Vera, Daniel Bolliger, Lorena Rossi, Elisa Felici, Mirko Di Rosa, Michiel de Jong, and Susy Paolini. "Usability and acceptance of the embodied conversational agent Anne by people with dementia and their caregivers: exploratory study in home environment settings." JMIR mHealth and uHealth 9, no. 6 (2021): e25891. https://doi.org/10.2196/25891

[14] Khoo, Weslie, Long-Jing Hsu, Kyrie Jig Amon, Pranav Vijay Chakilam, Wei-Chu Chen, Zachary Kaufman, Agness Lungu et al. "Spill the tea: When robot conversation agents support well-being for older adults." In Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction, pp. 178-182. 2023. https://doi.org/10.1145/3568294.3580067

[15] M. Iwamoto, N. Kuwahara, and K. Morimoto, “Comparison of Burdenon Youth in Communicating with Elderly using Images VersusPhotographs,” International Journal of Advanced Computer Science andApplications, vol. 6, no. 10, 2015. https://dx.doi.org/10.14569/IJACSA.2015.061023

[16] Z. Xiaochun, M. Iwamoto, and N. Kuwahara, “Evaluation of PhotoContents of Conversation Support System with Protocol AnalysisMethod,” International Journal of Advanced Computer Science andApplications, vol. 9, no. 4, 2018. https://dx.doi.org/10.14569/IJACSA.2018.090404

[17] Z. Xiaochun, C. Dong-Eun, P. Siriaraya, and N. Kuwahara, “SentimentAnalysis and Classification of Photos for 2-Generation Conversation in China,” International Journal of Advanced Computer Science andApplications, vol. 10, no. 10, 2019. https://dx.doi.org/10.14569/IJACSA.2019.0101061

[18] Lei Jiang, Panote Siriaraya, Dongeun Choi and Noriaki Kuwahara, "A Library of Old Photos Supporting Conversation of Two Generations Serving Reminiscence Therapy", Front. Psychol. 12:704236. doi: 10.3389/fpsyg.2021.704236

[19] Wang, Xin, Juan Li, Tianyi Liang, Wordh Ul Hasan, Kimia Tuz Zaman, Yang Du, Bo Xie, and Cui Tao. "Promoting Personalized Reminiscence Among Older Adults Through an AI-Driven Interactive Multimodal Photo Album", JMIR Aging 2024;7:e49415, http://dx.doi.org/10.2196/preprints.49415

[20] Carós, M., Garolera, M., Radeva, P., & Giro-i-Nieto, X. (2020, June). Automatic reminiscence therapy for dementia. In Proceedings of the 2020 International Conference on Multimedia Retrieval (pp. 383-387). https://doi.org/10.1145/3372278.3391927

[21] Pou-Prom, C., Raimondo, S., & Rudzicz, F. (2020). A conversational robot for older adults with Alzheimer’s disease. ACM Transactions on Human-Robot Interaction (THRI), 9(3), 1-25. https://doi.org/10.1145/3380785

[22] A. Algherairy, M. Ahmed, “A review of dialogue systems: current trends and future directions”, Neural Comput & Applic 36, pp. 6325–6351, 2024. https://doi.org/10.1007/s00521-023-09322-1

[23] Naim Zierau, Christian Hildebrand, Anouk Bergner, Francesc Busquet, Anuschka Schmitt and Jan Marco Leimeister, "Voice bots on the frontline: Voice-based interfaces enhance flow-like consumer experiences & boost service outcomes" Journal of the Academy of Marketing Science (2023) 51:823–842. https://doi.org/10.1007/s11747-022-00868-5

[24] Brett Christenson, Christine Ringler, and Nancy J. Sirianni, “Speaking fast and slow: How speech rate of digital assistants affects likelihood to use”, Journal of Business Research, Volume 163, 2023, 113907, https://doi.org/10.1016/j.jbusres.2023.113907.

[25] Jeung, J. L., & Huang, J. Y. C. (2024, May). Unlocking Memories with AI: Exploring the Role of AI-Generated Cues in Personal Reminiscing. In Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (pp. 1-6). https://doi.org/10.1145/3613905.3650979