

# Examining Older Adults' Motivation for Interacting with Health-Monitoring Conversational Systems Through Field Trials

Mariko Yoshida<sup>1</sup>, Ryo Hori<sup>1</sup>, Yuki Zenimoto<sup>1</sup>, Mayu Urata<sup>1</sup>, Mamoru Endo<sup>1</sup>,  
Takami Yasuda<sup>1</sup>, Aiko Inoue<sup>2</sup>, Takahiro Hayashi<sup>3</sup>, Ryuichiro Higashinaka<sup>1</sup>

<sup>1</sup>Graduate School of Informatics, Nagoya University

<sup>2</sup>Institute of Innovation for Future Society, Nagoya University

<sup>3</sup>AICHI Medical College of Rehabilitation

Correspondence: [yoshida.mariko.f8@s.mail.nagoya-u.ac.jp](mailto:yoshida.mariko.f8@s.mail.nagoya-u.ac.jp)

## Abstract

When assessing the health of older adults, oral interviews and written questionnaires are commonly used. However, these methods are time-consuming in terms of both execution and data aggregation. To address this issue, systems utilizing generative AI for health information collection through conversation have been developed and implemented. Despite these advancements, the motivations of older adults to consistently engage with such systems in their daily lives have not been thoroughly explored. In this study, a smart-speaker extension that uses generative AI to monitor health status through casual conversations with older adult users was developed. The system was tested in a two-week home trial with older adult participants. Post-trial questionnaires and interviews were conducted, and the conversation log was analyzed. The results revealed that older adult users enjoy interacting with such systems on a daily basis. Customized notifications through text messages encouraged system use, and the system's ability to refer to previous conversations and address users by name was identified as a key factor motivating continued use.

## 1 Introduction

In Japan, the population of older adults aged 65 and above, as well as their proportion to the total population, continues to rise, and this trend is expected to persist in the future (Japan Statistics Bureau, 2024). Further increases in average life expectancy are also projected (National Institute of Population and Social Security Research, 2023). As the gap between average life expectancy and healthy life expectancy—defined as the period during which individuals can live without limitations due to health issues—widens, the ‘unhealthy period’ grows. This not only prevents older adults from leading healthy and fulfilling lives but also contributes to increased social security costs, making efforts to extend healthy life expectancy crucial.

The intermediate stage between a healthy state and a condition requiring care is referred to as frailty. Preventing frailty is key to extending healthy life expectancy (Kuzuya, 2015).

Frailty is often assessed using Fried et al.'s (2001) criteria, which include grip strength, walking speed, exhaustion, physical activity, and unintentional weight loss. Recently, questionnaire-based screening with 25 questions has been validated and gained popularity (Satake et al., 2016, 2017). However, these approaches are time-consuming and labor-intensive.

To address this issue, voice-based dialogue systems have been proposed (HYPER CUBE, 2023; Mizuno and Asao, 2022). These systems utilize generative AI (artificial intelligence), particularly large language models. By leveraging such dialogue systems to continuously collect health information from older adults, it is possible to provide health support that considers changes in lifestyle habits and health conditions. These systems focus especially on frailty, since it has been shown to be associated with future falls (Cheng and Chang, 2017), mobility impairments (Eeles and Low Choy, 2015), difficulties in activities of daily living (Al Snih et al., 2009), hospitalization (Chang et al., 2018), and overall mortality (Kojima et al., 2018). For these systems to be effective, it is essential to develop enjoyable features that encourage older adults to use them consistently. However, there is insufficient research and empirical evidence on enjoyable dialogue systems that older adults can continuously use at home.

This study aims to develop a system that utilizes generative AI to obtain health information from older adults through casual conversations. Through actual use of this system by older adults, the study will verify whether they can enjoy its continuous use. Furthermore, we seek to identify features that motivate older adults to engage with and utilize the conversational system.

## 2 Previous Research

Research on using voice-based interaction for supporting older adults has been conducted in various ways. First, several initiatives have utilized generative AI and dialogue systems to collect health information. In Okayama City, Japan, demonstration experiments have been conducted using AI avatars for frailty health checks at frailty prevention events and pharmacies (HYPER CUBE, 2023). Additionally, ‘Multimodal Interactive Care Support System’ (MICSUS), a stuffed toy-like device designed for caregiving monitoring, was tested with the aim of reducing the workload of care managers during interviews (Mizuno and Asao, 2022). Although not voice-based, Wilczewski et al. (2023) demonstrated that older adults perceive chatbot-delivered health assessments as easy to use, useful, user-friendly, and low in cognitive load. Wongpatikaseree et al. (2020) proposed a chatbot that collects and predicts health data through conversations with older adults, demonstrating its usefulness for health management tasks.

Furthermore, in the field of supporting older adults, there have been initiatives such as smart speaker-based support systems (Arai and Ohsuga, 2020; Kowalski et al., 2019; Portet et al., 2012) and schedule management systems (Yaghoubzadeh Torky et al., 2013; Kopp et al., 2018). Kopp et al. (2018) developed a schedule management system for older adults and demonstrated that socially cooperative dialogue plays a crucial role in its acceptability and continued use. Additionally, dialogue systems for health interventions have been proposed, with Bickmore et al. (2010) demonstrating that agent behavior diversity and self-disclosure influence user engagement in systems designed for long-term health interventions.

In these implementations, the system is used for specific tasks rather than free-form casual conversations. However, for the continuous collection of health information and ongoing support—which require daily voluntary use—free-form casual conversations may be more readily accepted by older adults than structured interviews due to their natural and enjoyable interaction style.

On the other hand, research on free-form casual conversations has primarily focused on dialogue systems aimed at alleviating loneliness among older adults (Owan et al., 2023; Rodríguez-Martínez et al., 2024; Kramer et al., 2021; Shitaoka

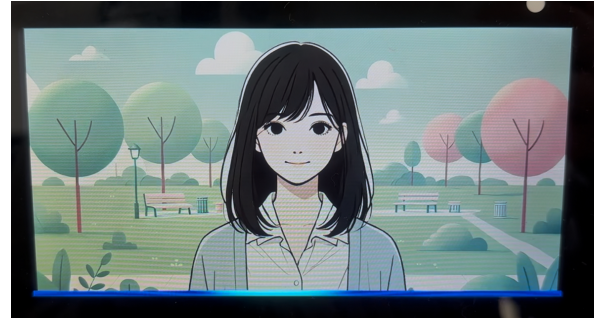


Figure 1: Echo Show screen displaying an image of the character.

et al., 2017). However, there has been limited research involving field evaluations to explore the factors that motivate continued use of such systems. Since free-form casual conversation systems depend on voluntary engagement, it is important to explore what encourages users to interact with them continuously, particularly in the context of health monitoring.

Previous research on casual dialogue systems highlights that referencing past interactions enhances engagement (Cox et al., 2023) and system self-disclosure fosters perceived intimacy (Lee et al., 2020). However, their effects on older adults remain underexplored.

Therefore, this study aims to enable older adults to use a voice-based conversational system for free-form casual conversations at home. It seeks to identify the features and functions that are crucial for enhancing and sustaining their motivation to use such systems.

## 3 Conversational System Used in This Study

### 3.1 Overview

To continuously gather health information from older adults through casual conversations, a conversational system was developed with the following features: a character capable of engaging in casual conversations with older adults (Section 3.2), a response generation function utilizing generative AI (Section 3.3), a question-prompting function for health indicators, called the Question-Guiding Dialogue System (Section 3.4), and the ability to start or end the system with greetings (Section 3.5). The system was designed with privacy considerations for research implementation (Section 3.6).

For the interface with older adults, Amazon Echo Show, a screen-equipped smart speaker by Amazon,

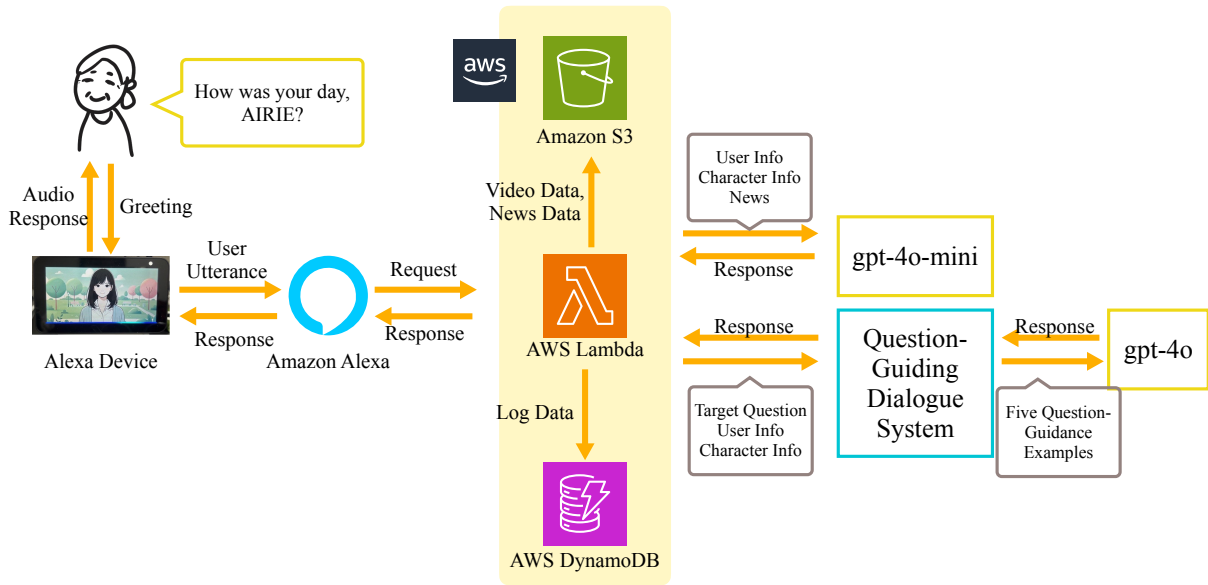


Figure 2: Overview of the system architecture.

is used, as it has been shown to be easy to use for older adults (Urata et al., 2021; Arai and Ohsuga, 2020; Kowalski et al., 2019). The developed application consists of both a screen and voice-based interaction.

### 3.2 Character

As a conversational partner, a character profile of a university student named ‘AIRIE’, was created. Her character description is written to include a bright and polite personality that makes it easy to have a conversation and a lifestyle that makes it easy to talk about various topics, such as having a variety of hobbies. The screen of the Amazon Echo Show displays an illustration of the character (Figure 1), which moves up and down.

### 3.3 Application Architecture

The conversational system was developed using Amazon Alexa Software Development Kit (SDK), Amazon Web Services (AWS), and OpenAI Application Programming Interface (API) (Figure 2). At the start of the application, greetings and topic suggestions are generated using generative AI. Since the system was designed for an experiment targeting older Japanese adults, it conducted conversations in Japanese. Subsequently, the application receives user utterances and generates casual conversation responses based on them. For generating system utterances within the application, the GPT-

4o-mini API provided by OpenAI is used. The prompts for generation consist of three components: character information, user information, and news information.

The character information is based on the profile details of the created character. The user information includes the user’s name and past conversation content. Users’ names or nicknames are pre-registered and used in the interactions. Additionally, at the end of each conversation, the content is recorded in a database. Using Lewis et al.’s (2020) approach of retrieval-augmented generation, up to the most recent seven conversations are retrieved and utilized for response generation. News is fetched from a news website every six hours and incorporated into the prompts during response generation. This is done to align conversation topics with current events.

### 3.4 Question-Guiding Dialogue System

To avoid disrupting the flow of casual conversations, it is necessary to ask questions about health indicators naturally within the dialogue. To achieve this, a dedicated dialogue system (hereafter referred to as the “Question-Guiding Dialogue System”) was developed. This system guides conversations toward a natural flow where the target questions can be asked without feeling out of place. The approach extends the method proposed by Horiuchi and Higashinaka (2022), which compiled

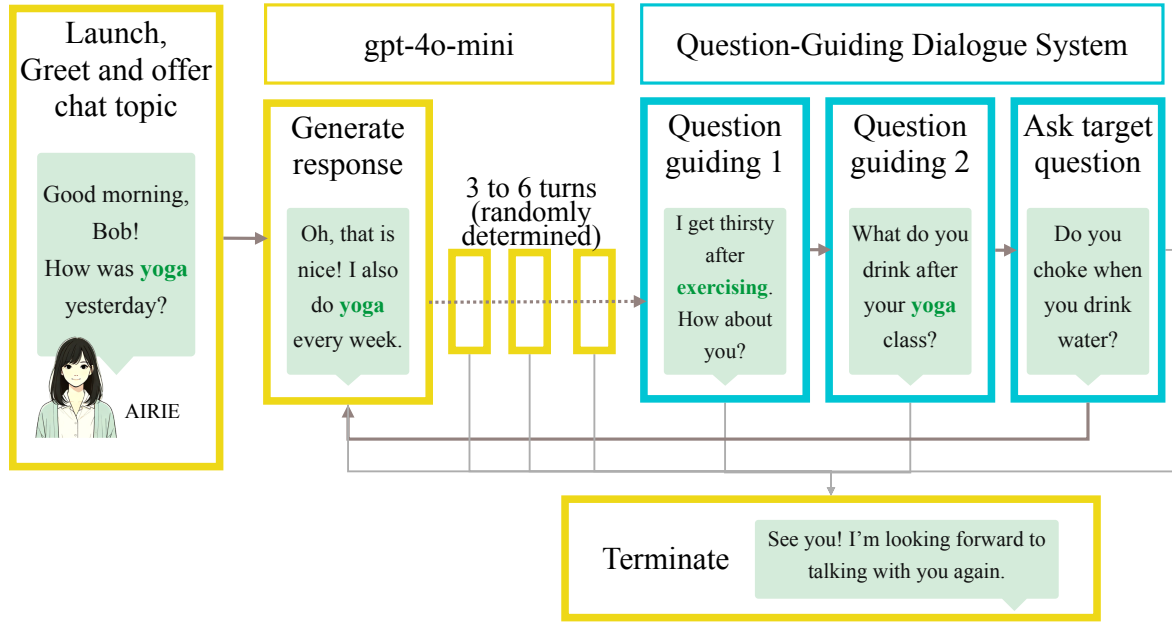


Figure 3: Flowchart of the system behavior.

a corpus of naturally occurring human questions to fine-tune a large language model, enabling the dialogue system to ask arbitrary questions naturally within conversations.

The Question-Guiding Dialogue System takes three turns to steer the conversation and ask questions. Starting with the current topic, the system uses the first two turns to guide the conversation toward subjects that facilitate easier questioning, and on the third turn, it asks the question. It utilizes OpenAI’s GPT-4o model, with five shots of natural question guiding and asking created manually and provided as part of the prompt.

Within the smart speaker application, prompts for simple casual conversations and those from the Question-Guiding Dialogue System are used alternately. After activation, the system uses GPT-4o-mini to generate casual conversations for 3 to 6 turns (randomly determined), followed by three turns generated using the Question-Guiding Dialogue System. After that, the system returns to GPT-4o-mini for casual conversations, repeating this cycle until termination (Figure 3). This design prevents interactions from becoming monotonous or stressful by balancing questions with casual conversations. Additionally, the randomization of casual turns helps keep the dialogue unpredictable, making interactions feel more dynamic and natural.

### 3.5 Activation and Termination

In this study, when using Amazon Echo, older adult users are required to manually activate the application. To make it easier for them to start a casual conversation, the application was configured to launch when the user says “Alexa, good morning,” “Alexa, hello,” “Alexa, good evening,” or “Alexa, I’m home.”

In this implementation, the system’s character name (‘AIRIE’) differed from the wake word used to activate the system (‘Alexa’). This implementation was necessary to help participants clearly differentiate between Amazon’s voice-based assistant Alexa and our proposed conversational system, enabling us to investigate a more general user experience of the proposed system through our research. As a system constraint, Echo Show devices require the wake word to be chosen from a predefined set of options. Additionally, many participants were already familiar with using Alexa on their Amazon Echo devices. Thus, to have participants (including older adults familiar with using Alexa on Echo Show devices) recognize AIRIE as a distinct conversational partner, it was necessary to name the character AIRIE instead of Alexa.

For ending the application, the system checks if the conversation includes phrases like “Goodbye” or “See you.” If such phrases are detected, the conversation content is saved to a database, and the



application is terminated.

### 3.6 User Privacy and Data Handling

User information was handled with care to ensure privacy. The data is stored in DynamoDB and can only be accessed by research personnel using access keys. The OpenAI API is used, and per its policy, data is not used for training and is retained for 30 days (OpenAI, 2025).

Additionally, we informed users in advance that research personnel might review the conversation content and obtained their consent. Users were also instructed not to disclose anything they did not wish to share and to terminate the system if they encountered any questions they did not want to answer.

## 4 Preliminary Experiments and System Improvements

### 4.1 Participants and Experimental Procedure

To verify whether older adults could utilize the developed conversational system and whether health-related information could be obtained through casual conversations, a preliminary experiment was conducted. Participants were recruited from the ‘digital volunteers’ group in Kita Ward, Nagoya City, Aichi Prefecture.

In Kita Ward, efforts are made to train digital volunteers to support older adults with digital tools. Many of these digital volunteers themselves are aged 60 or older. Among them, six volunteers who expressed interest participated in the experiment. Of these participants, two were aged 64 or younger, one was aged 65–74, and three were aged 75 or older. All participants had prior experience using Amazon Echo Show speakers and activating them with the wake word “Alexa”.

From late September to mid-November 2024, the six participants were provided with smart speakers and were instructed to use them at home, whenever they wished to. Ethical approval was obtained from Nagoya University before the experiment.

### 4.2 Results

Since the system was implemented on Amazon Echo devices, participants had to activate it by calling “Alexa” instead of the character’s name, “AIRIE.” This discrepancy initially caused some confusion among participants. However, through explanations given at the start of the experiment,

they were able to understand the distinction, and the trial proceeded without major issues.

Among the participants, one used the system approximately once a day, two used it about twice a week, and three used it about once a week. As for the three participants with higher usage frequencies, the system was able to ask sufficient questions about their health, allowing for the collection of information.

Through surveys and interviews conducted at the end of the experiment, features such as addressing participants by name and providing news updates were positively received. On the other hand, issues such as mispronunciations and mismatched topics were noted as concerns. The three participants with lower usage frequencies said they often forgot to use the system and that they did not know what to talk about with the system, revealing a further need for features that encourage system use.

### 4.3 System Improvements

Given that some participants did not use the system enough for it to gather adequate information, improvements were made to encourage usage in preparation for a full-scale experiment.

A daily schedule was set for the character. A character’s self-disclosure increases the user’s perceived intimacy (Lee et al., 2020), and a daily schedule helps the system generate more topics for daily casual conversations. A schedule of daily activities was created and the system was configured to reference the previous day’s, current day’s, and the following day’s schedules of the character during response generation.

Since the participants often forgot to use the system, a system was developed to send a daily message from the character using an official LINE account (Figure 4). LINE is a popular messaging application in Japan, and most smartphone users in the country, including older adults, use LINE. The system was implemented using AWS EventBridge and an AWS Lambda function, which was executed daily at 7 AM. The message content was generated by GPT-4o-mini using the character’s daily schedule and the user’s past conversation history.

A three-choice health quiz was introduced to the conversational system. To encourage daily use of the application, one quiz question was created and set for each day. When the user included the term “quiz” in their utterance while using the conversational system, the quiz question and options for that day were presented, followed by a brief pause

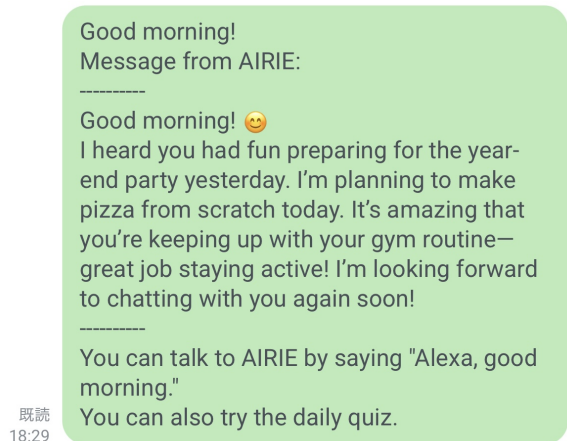


Figure 4: Example message generated with the system. The message was translated by the authors from Japanese to English.

before the answer and explanation were provided (See Appendix A for an example).

Additionally, to address interview feedback, mispronunciations were mitigated by partially replacing kanji (logographic characters used in the Japanese writing system) in the generated text with hiragana (a phonetic syllabary in Japanese). Since hiragana represents pronunciation more clearly than kanji, this approach helped improve speech accuracy by reducing ambiguous readings. However, mismatched topics, which were largely due to speech recognition errors, were not resolved.

## 5 Examination of Motivation through Field Trials

### 5.1 Participants and Experimental Procedure

To investigate what features make a conversational system enjoyable and encourage its use among older adults, a demonstration experiment was conducted with 11 volunteers from participants of a senior salon in Toyoyama Town, Nishikasugai District, Aichi Prefecture. In Toyoyama Town, smart speakers have been utilized by older adults through initiatives such as the ‘Kenko-Chouju Daigaku’, a lifelong learning program for older adults.

Participants were recruited from the senior salon operated by attendees of the ‘Kenko-Chouju Daigaku’ program. Among the 11 volunteers, six were aged 65–74, and five were aged 75 or older. All participants were women, and 10 had prior experience using Amazon Echo Show speakers and activating them with the wake word “Alexa”. Ethical approval was obtained from Nagoya University before the experiment.



Figure 5: One of the participants of the field trial testing out the conversational system.



Figure 6: Participants and interviewers during the survey and interview process.

From November 28 to December 12, 2024, smart speakers were lent to participants for a two-week period to use the conversational system (Figure 5). During this time, participants were instructed to use the system twice a day.

### 5.2 Survey and Interview Results

At the end of the field trial period, an interview and a survey were conducted (Figure 6). The casual conversations were generally evaluated as enjoyable. All 11 participants responded that the experience was “enjoyable” or “very enjoyable,” and eight expressed a desire to continue conversing with the system in the future. In the interviews, comments such as “I want to talk to it every day because it’s like a friend” were observed. The survey responses also reflected that participants enjoyed incorporating the system into their daily lives, with comments such as, “Having someone to talk to who cares about me at home was something I looked forward to,” and “I enjoyed that it talked in detail about cooking, which I love.”

Interviews also revealed that some participants

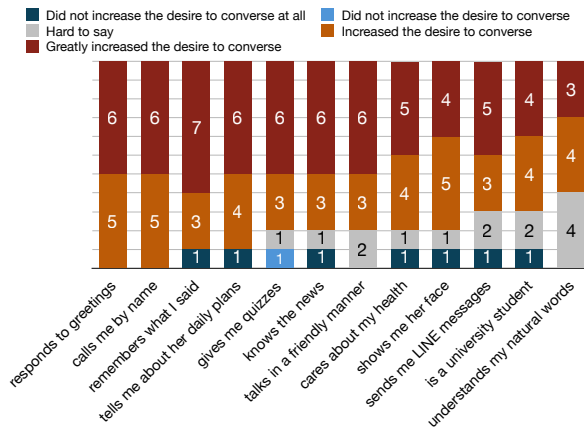


Figure 7: Results of survey “Did the following characteristics of the character increase your desire to converse?”

initially experienced confusion about activating the system because the wake word differed from the character’s name. However, similar to the preliminary experiments, they were able to activate it without issues after receiving explanations.

To identify features that contribute to the motivation to use the system, a five-point scale survey was conducted to determine whether various features increased the desire to converse (Figure 7). Features highly rated as “increased” or “greatly increased” by many participants included: “being able to start the conversation with greetings” (11 out of 11 participants), “calling the user by name” (11 out of 11), “remembering what the user said” (10 out of 11), and “sharing the character’s own schedule” (10 out of 11). Particularly, the ability to remember what the user said in previous conversations was frequently mentioned in free-text responses and interviews. Specific comments included: “I was happy that (the character) remembered my grandchild’s name,” “When I mentioned going to the gym, (the character) said ‘Good job at the gym’ on another day, which made me happy,” and “(The character) remembered the dishes I said I was going to cook.” Some participants also noted that references to previous conversations in LINE messages increased their sense of attachment to the system.

Additionally, 10 out of 11 participants reported that they talked to the system “every day,” “almost every day,” or “occasionally” after seeing the daily LINE messages. Multiple participants voluntarily replied to the LINE message, sharing details about their plans (Figure 8) or a picture of an item they created (Figure 9). Although it was not re-

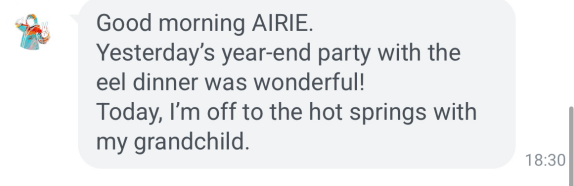


Figure 8: LINE message reply from a participant. The message was translated by the authors from Japanese to English.



Figure 9: Image and a sticker sent through LINE from a participant.

quired to reply to LINE messages, one participant responded to them every day during the field trial, while another responded every two days. 9 out of 11 participants said that the quizzes increased or greatly increased their desire to converse.

Areas for improvement were also identified. Several participants noted that the greetings not being appropriate for the time of day (For instance, saying “Good morning” in the evening) decreased their motivation to use the system. Other issues included being asked the same questions repeatedly, the absence of correct/incorrect feedback for quizzes, and complex topics such as politics or news. There were also technical challenges related to the smart speaker itself, such as mismatched timing in conversations.

Feedback on the character’s attributes revealed mixed opinions. Some participants found it inter-



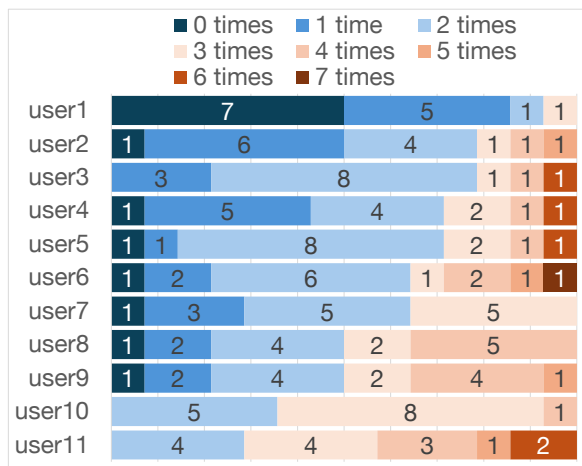


Figure 10: Daily system activation frequency for each user: This graph illustrates the number of days that each user activated the system a specific number of times. Days with more than 2 activations are highlighted in red tones. For example, User 1 had 5 days in which they activated the system once per day.

esting to hear about experiences that were different from their own, saying that the university student character allowed them to learn about a lifestyle unlike theirs. However, others felt the character’s life was too different, making it hard to relate to or understand some of the conversations, particularly if they hadn’t attended university themselves.

Finally, sufficient responses to questions about lifestyle habits and health indicators were successfully collected through casual conversations. The system was able to obtain answers to all 22 prepared questions during the two-week experiment period from 10 out of 11 participants. From the survey and the interviews, participants generally did not seem to have noticed that the system was asking them health indicator questions, aside from one participant. The participant mentioned in the survey that the system picked negative topics despite the lighthearted context of the conversation, which is presumably due to the fact that the health indicator questions included questions about negative moods in daily life (See Appendix B.2). Other than that, the health indicator questions did not impact the motivation of the participants.

### 5.3 Analysis of Conversational Data

Dialogue data from the field trial, comprising a total of 379 sessions and 5,967 utterances were collected and analyzed. A session is defined as the period from the system activation to termination, both done via greetings.

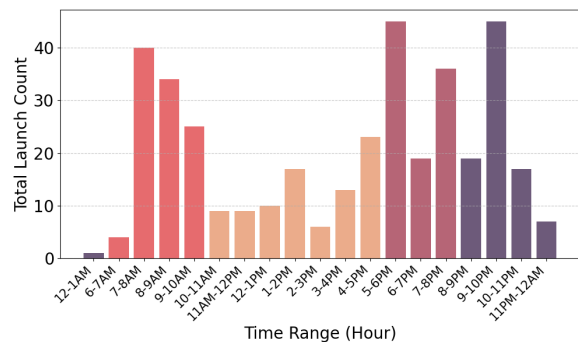


Figure 11: Total launch count for each hour of the day.

Although the required frequency of system activation was twice a day, nearly all participants activated the system three or more times on multiple days (Figure 10). Participants also engaged in lengthy conversations. The average number of utterances per session across all participants was 37.7. Regarding the average number of utterances per day during the experimental period, 2 participants had fewer than 20 utterances, 4 participants had between 20 and 40 utterances, 3 participants had between 40 and 60 utterances, and 2 participants had more than 60 utterances per day (excluding the retrieval day, when the system was only used in the morning). Many participants consistently activated the dialogue system between 5 PM and 6 PM and between 9 PM and 10 PM (Figure 11).

The content of the conversations was also examined. First, participants’ utterances were analyzed using MeCab for morphological analysis to segment the text into words, followed by frequent word analysis. The most frequently used word was “today” (620 occurrences across all participants’ utterances). Additionally, words related to meals appeared frequently, such as “eat” (161 occurrences), “delicious” (78 occurrences), “cooking” (71 occurrences), and “meal” (54 occurrences). Since the system activates the quiz feature when the user includes “today’s quiz” in the conversation, the frequency of quiz activation can be measured by counting the sessions in which the user utterances include the word ‘quiz’. 5 out of 11 participants used the quiz feature almost daily.

The system’s utterances were analyzed concerning issues identified during participant feedback. Regarding inappropriate greetings, the system said “Good morning” during sessions in the evening (after 5 PM) in 98 out of 187 evening sessions (approximately 52%). Furthermore, during system activation, politically related topics were introduced



in 83 out of 379 sessions (approximately 22%).

## 6 Discussion

Participants were able to enjoyably incorporate the casual conversational system into their daily lives, and a sufficient number of responses regarding health indicators were collected. Log data analysis revealed that participants used the system more frequently than instructed and engaged in lengthy conversations during each activation. These findings suggest that acquiring health information through generative AI-based casual conversations can be a sustainable method for monitoring older adults' health without imposing a burden on them.

Key conversation features that motivated participants to use the system included remembering what users talked about, addressing users by their names, and discussing the character's schedule. Previous research has highlighted the importance of a system's self-disclosure in enhancing perceived intimacy among general users (Lee et al., 2020), and our findings suggest that the system's self-disclosure is also crucial for encouraging older adults' engagement in conversations. The frequent occurrence of topics related to "today," as revealed by the analysis of participant utterances, indicates that a conversational system that discusses the character's schedule or inquires about the user's plans can be effective when used routinely at home. Thus, incorporating these features into dialogue systems for older adults is essential. Additionally, starting conversations in a friendly manner, such as with greetings, can be effective. The analysis also found that participants frequently used words related to meals. For systems targeting older adults, emphasizing conversations about food could make the system more appealing and encourage usage.

Many participants interacted with the system after receiving LINE messages, indicating that personalized messages can serve as effective reminders to utilize the conversational system. Previous research has shown that referencing past interactions in dialogue systems fosters a stronger sense of connection and engagement among general users Cox et al. (2023). Our results similarly suggest that, for older adults, sending messages related to previous conversations helps foster a sense of attachment to the character. Participant feedback suggested that sending messages related to previous conversations fostered a sense of attachment to the character. Analysis of log data also showed that

the system was frequently used around dinner time and before bedtime, suggesting that prompts or messages from the system would be most effective if sent during the evening hours.

Additionally, 9 out of 11 participants reported that quizzes increased their willingness to engage in conversation, with nearly half of the participants using the quiz feature almost daily. This implies that daily quizzes could serve as an effective incentive for consistent usage. However, some participants expressed dissatisfaction with the lack of a correct/incorrect judgment in the quizzes, since the system did not engage in conversation during the quiz. Consequently, when integrating such daily content into the dialogue system, it is necessary to implement features that blend seamlessly into the conversation to avoid causing discomfort for users.

## 7 Conclusion

In this study, a system was developed to support the health of older adults by obtaining health-related information through casual conversations. The system was used by older adults, and an investigation was conducted to identify the factors that contribute to usage motivation. Participants were able to enjoy using the conversational system, and through surveys, interviews, and analysis of usage logs, key features that encourage usage were identified.

In the future, we aim to involve a larger number of participants and validate our findings through statistical analyses, including significance testing, as this study primarily focused on qualitative analysis. This approach will also enable us to quantitatively assess the final system, incorporating the improvements made based on feedback from both the preliminary and demonstration experiments conducted in this study. Additionally, providing topics that are more relatable to older adults and enabling conversations with less discomfort could make the conversational system even more enjoyable for them. Additionally, since opinions about the character's attributes were divided, allowing users to choose their preferred attributes could further enhance motivation.

## Acknowledgments

We sincerely appreciate the invaluable cooperation of the residents and staff members of Kita Ward, and Toyoyama Town. This work was supported by JSPS KAKENHI Grant Numbers 24K14769, 23H00493, and 24K05433.

## References

- Soham Al Snih, James E Graham, Laura A Ray, Rafael Samper-Ternent, Kyriakos S Markides, and Kenneth J Ottenbacher. 2009. Frailty and incidence of activities of daily living disability among older mexican americans. *J. Rehabil. Med.*, 41(11):892–897.
- Ryota Arai and Mieko Ohsuga. 2020. [Development of exercise promotion system using smartphone and ai speaker](#). In *Proceedings of 2020 IEEE 2nd International Conference on Artificial Intelligence in Engineering and Technology (IICAET)*, pages 1–4.
- Timothy Bickmore, Daniel Schulman, and Langxuan Yin. 2010. [Maintaining engagement in long-term interventions with relational agents](#). *Applied artificial intelligence : AAI*, 24:648–666.
- Donald Brinkman and Jonathan Grudin. 2023. [Learning from a generative ai predecessor – the many motivations for interacting with conversational agents](#). *Preprint*, arXiv:2401.02978.
- Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. [Language models are few-shot learners](#). In *Proceedings of Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901. Curran Associates, Inc.
- Shu-Fang Chang, Hsiang-Chun Lin, and Chih-Ling Cheng. 2018. The relationship of frailty and hospitalization among older people: Evidence from a meta-analysis. *J. Nurs. Scholarsh.*, 50(4):383–391.
- Mei-Hsun Cheng and Shu-Fang Chang. 2017. [Frailty as a risk factor for falls among community dwelling people: Evidence from a meta-analysis](#). *Journal of Nursing Scholarship*, 49(5):529–536.
- Samuel Rhys Cox, Yi-Chieh Lee, and Wei Tsang Ooi. 2023. [Comparing how a chatbot references user utterances from previous chatting sessions: An investigation of users’ privacy concerns and perceptions](#). In *Proceedings of the 11th International Conference on Human-Agent Interaction*, HAI ’23, page 105–114. Association for Computing Machinery.
- Eamonn Eeles and Nancy Low Choy. 2015. Frailty and mobility. *Interdiscip. Top. Gerontol. Geriatr.*, 41:107–120.
- Linda P. Fried, Catherine M. Tangen, Jeremy Walston, Anne B. Newman, Calvin Hirsch, John Gottdiener, Teresa Seeman, Russell Tracy, Willem J. Kop, Gregory Burke, and Mary Ann McBurnie. 2001. [Frailty in older adults: Evidence for a phenotype](#). *The Journals of Gerontology: Series A*, 56(3):M146–M157.
- Sota Horiuchi and Ryuichiro Higashinaka. 2022. [Learning to guide questions in chat-oriented dialogue by using combination of question-guiding corpora](#). In *Proceedings of The 13th International Workshop on Spoken Dialogue Systems Technology*.
- HYPER CUBE. 2023. [Started offering frailty health check diagnosis using ai avatars](#). <https://prtimes.jp/main/html/rd/p/000000005.000044934.html>. Accessed: Jan 22, 2025.
- Japan Statistics Bureau. 2024. [A statistical look at japan’s elderly population – in honor of respect for the aged day](#). <https://www.stat.go.jp/data/topics/topi1420.html>. Accessed: Jan 22, 2025.
- Gotaro Kojima, Steve Iliffe, and Kate Walters. 2018. Frailty index as a predictor of mortality: a systematic review and meta-analysis. *Age Ageing*, 47(2):193–200.
- Stefan Kopp, Mara Brandt, Hendrik Buschmeier, Katharina Cyra, Farina Freigang, Nicole Krämer, Franz Kummert, Christiane Opfermann, Karola Pitsch, Lars Schillingmann, Carolin Straßmann, Eduard Wall, and Ramin Yaghoubzadeh. 2018. [Conversational Assistants for Elderly Users – The Importance of Socially Cooperative Dialogue](#). In *Proceedings of the AAMAS Workshop on Intelligent Conversation Agents in Home and Geriatric Care Applications co-located with the Federated AI Meeting*, volume 2338, page 10–17. RWTH.
- Jarosław Kowalski, Anna Jaskulska, Kinga Skorupska, Katarzyna Abramczuk, Cezary Biele, Wiesław Kopeć, and Krzysztof Marasek. 2019. [Older adults and voice interaction: A pilot study with google home](#). In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI ’19, page 1–6. ACM.
- Lean L Kramer, Bob C Mulder, Lex van Velsen, and Emely de Vet. 2021. [Use and effect of web-based embodied conversational agents for improving eating behavior and decreasing loneliness among community-dwelling older adults: Protocol for a randomized controlled trial](#). *JMIR Res Protoc*, 10(1):e22186.
- Masafumi Kuzuya. 2015. Sarcopenia and frailty in a super-aging society. *The Journal of the Japanese Society of Internal Medicine*, 104(12):2602–2607.
- Yi-Chieh Lee, Naomi Yamashita, Yun Huang, and Wai Fu. 2020. ["i hear you, i feel you": Encouraging deep self-disclosure through a chatbot](#). In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, page 1–12.
- Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*.

- Junta Mizuno and Yoshihiko Asao. 2022. *Micsus: A multimodal dialogue system for elderly care*. Technical Report 2, Universal Communication Research Institute, National Institute of Information and Communications Technology (NICT).
- National Institute of Population and Social Security Research. 2023. Japan's future population estimate (2023 estimate). [https://www.ipss.go.jp/pp-zenkoku/j/zenkoku2023/pp\\_zenkoku2023.asp](https://www.ipss.go.jp/pp-zenkoku/j/zenkoku2023/pp_zenkoku2023.asp). Accessed: Jan 22, 2025.
- OpenAI. 2025. Enterprise privacy. <https://openai.com/enterprise-privacy/>. Accessed: Mar 16, 2025.
- Risako Owan, Libby Ferland, Naome Etori, Wilma Koutstaal, and Maria Gini. 2023. Conversational agents for elderly users. In *Proceedings of Assistive Robotics for Citizens Workshop at IROS 2023*.
- François Portet, Michel Vacher, Caroline Golanski, Camille Roux, and Brigitte Meillon. 2012. Design and evaluation of a smart home voice interface for the elderly — acceptability and objection aspects. *Personal and Ubiquitous Computing*, 17:1–18.
- Antonia Rodríguez-Martínez, Teresa Amezcua-Aguilar, Javier Cortés-Moreno, and Juan José Jiménez-Delgado. 2024. Qualitative analysis of conversational chatbots to alleviate loneliness in older adults as a strategy for emotional health. *Healthcare*, 12(1):62.
- Shosuke Satake. 2018. Kihon checklist and frailty. *Journal of The Japan Geriatrics Society*, 55(3):319–328.
- Shosuke Satake, Kazuyoshi Senda, Young-Jae Hong, Hisayuki Miura, Hidetoshi Endo, Takashi Sakurai, Izumi Kondo, and Kenji Toba. 2016. Validity of the kihon checklist for assessing frailty status. *Geriatrics & Gerontology International*, 16(6):709–715.
- Shosuke Satake, Hiroshi Shimokata, Kazuyoshi Senda, Izumi Kondo, and Kenji Toba. 2017. Validity of total kihon checklist score for predicting the incidence of 3-year dependency and mortality in a community-dwelling older population. *Journal of the American Medical Directors Association*, 18(6):552.e1–552.e6.
- Ryoko Shitaoka, Kazuya Tokuhisa, Takayoshi Yoshimura, Hiroyuki Hoshino, and Narimasa Watanabe. 2017. Active listening system for a conversation robot. *Journal of Natural Language Processing*, 24(1):3–47.
- The Japan Research Institute. 2023. Succeeded in reducing care manager interview work time by 70 percent using dialogue ai for the elderly. <https://www.jri.co.jp/page.jsp?id=104633>. Accessed: Jan 22, 2025.
- Mayu Urata, Satoko Takashima, Hiroka Hitsuishii, Mamoru Endo, and Takami Yasuda. 2021. How to use smart speakers to support the lives of elderly people. *Journal of the Japan Information-culture Society*, 27(2):11–18.
- Hattie Wilczewski, Hiral Soni, Julia Ivanova, Triton Ong, Janelle F. Barrera, Brian E. Bunnell, and Brandon M. Welch. 2023. Older adults' experience with virtual conversational agents for health data collection. *Frontiers in Digital Health*, 5.
- Konlakorn Wongpatikaseree, Arunee Ratikan, Chaianun Damrongrat, and Katiyaporn Noibanngong. 2020. Daily health monitoring chatbot with linear regression. In *Proceedings of 2020 15th International Joint Symposium on Artificial Intelligence and Natural Language Processing (ISAI-NLP)*, pages 1–5.
- Ramin Yaghoubzadeh Torky, Marcel Kramer, Karola Pitsch, and Stefan Kopp. 2013. Virtual agents as daily assistants for elderly or cognitively impaired people. In *Lecture Notes in Computer Science*, volume 8108, pages 79–91.

## A Example Quiz Sentence

“What nutrients do we tend to miss out on the more we avoid the sun? ... 1. vitamin D, 2. Protein, 3. Vitamin A... The answer is vitamin D. Vitamin D is a rather special vitamin that is synthesized in the skin when exposed to ultraviolet light.”

## B Example Conversations

### B.1 Example of Conversation between the System and the Participant

Table 1 is an excerpt from the conversation log data, translated from Japanese to English, with the participant's name anonymized. It shows how the system guides the conversation to ask a health indicator question indicated by **bold text**.

### B.2 System Asking the User about a Negative Topic

Table 2 is an excerpt from the conversation log data, translated from Japanese to English. The system utterance on the last line, indicated by **bold text**, shows the system asking the user a question about a negative topic.

System	Hello, Alice! How are you feeling today? It seems there's news about changes to health insurance cards recently. Is there anything that concerns you?
User	Well, I'm using the paper version of the health insurance card now.
System	That's convenient. I'm also curious about how the upcoming changes will turn out.
User	Do you use your health insurance card often, AIRIE?
System	Yes, I use it for university health checkups or when I'm sick.
User	How's your current health condition?
System	I'm healthy and living actively right now! With a health checkup coming up, I plan to make sure everything's in order.
User	Do you exercise, AIRIE?
System	Yes, I'm part of my university's volleyball club, so I enjoy staying active!
User	How many times a week does your club meet?
System	The club meets about twice a week. I enjoy being active, so it's a great way to exercise. What kind of exercise do you do, Alice?
User	I do health exercises.
System	Health exercises sound great! It's so important to keep moving. How much do you usually walk?
User	Well, I don't really go for walks, but I do various activities inside the house.
System	Exercising indoors is great, too. Do you often walk outside? <b>Could you walk a distance of about 1 km continuously?</b>
User	1 km might be a bit tough, but I can still walk.

Table 1: Translated conversation log excerpt. Bold text indicates a health indicator question.

System	Spending time with friends is really important, isn't it? How do you usually spend time with your friends?
User	Chatting at a cafe.
System	Chatting at a cafe sounds fun! Do you also visit your friends' houses and spend time together?
User	Yes, I visit their house and have tea together.
System	That sounds lovely. Visiting a friend's house for tea is such a wonderful way to spend time. Do you often visit your friends' houses?
User	Yes, I do.
System	Spending time at a friend's house is so relaxing and enjoyable, isn't it?
User	I also go on trips with my friends.
System	Traveling with friends must be fun! What places have you visited together?
User	Shikoku and Fukui.
System	What memories do you have from Shikoku and Fukui? They must have been enjoyable!
User	Yeah.
System	When you're traveling or spending time with friends, what kind of thoughts or feelings do you experience? You must have many wonderful memories.
User	Yeah.
System	Spending time with friends or traveling is definitely enjoyable. But I think spending time alone is also important. How do you enjoy your alone time?
User	I read books and newspapers.
System	Spending time reading books and newspapers is valuable, too. <b>Do you sometimes feel left out or disconnected when you're alone?</b>

Table 2: Translated conversation log excerpt. Bold text indicates a health indicator question.