Usability Evaluation of a Chatbot for Fitness and Health Recommendations Among Seniors in Assisted Healthcare

William Philipp[1], Ali Gölge[2], Andreas Hein[3] and Sebastian Fudickar[1]

[1]*Institute of Medical Informatics, University of Luebeck, Luebeck, Germany*

[2]*ATLAS Elektronik GmbH, Bremen, Germany*

[3]*Carl von Ossietzky University Oldenburg, Oldenburg, Germany*

{*w.philipp, sebastian.fudickar*}*@uni-luebeck.de, andreas.hein@uol.de, a.goelge@gmail.com*

Keywords: Senior Healthcare, Assisted Healthcare, Human-Computer Interaction (HCI) in Healthcare, Elderly Technology Use.

Abstract: This study explores the acceptance of seniors for a chatbot designed to support in maintaining activity levels and quality of life in an assisted healthcare setting. Building on findings from the TUMAL study, which developed a self-assessment tool for physical functioning, a proof-of-concept chatbot was created as an Android app. The chatbot enables users to view their health data, inquire about activity levels, and receive recommendations based on their results. A study involving 12 seniors (aged 75+) was conducted to evaluate the chatbot's usability and the participants' attitudes toward its recommendations. The System Usability Scale (SUS) revealed a suboptimal usability score of 66.3, with wide-ranging results indicating varying user experiences. While fitness-related recommendations were positively received, health-related advice prompted mostly negative feedback. Despite these challenges, the data querying functionality was considered useful, demonstrating a degree of acceptance among the senior user group. The study suggests that the participants' technical proficiency may have influenced their overall usability ratings.

1 INTRODUCTION

1.1 Motivation

The demographic shift poses one of the greatest challenges for industrialized nations. The World Health Organization (WHO) projects that by 2025, the number of people over 60 will rise to 1.5 billion (Röcker, 2012). This aging population will lead to a relevant increase in the demand for healthcare personnel in countries like Germany, which may not be met (Wolf et al., 2017). Currently, 15% of Europe's population reports difficulty performing daily tasks due to physical limitations, increasing the need for care. Chronic diseases and declining physical abilities are the main drivers of this demand (Röcker, 2012). To address these challenges, the German Federal Ministry of Education and Research has focused on "Ambient Assisted Living" (AAL) systems since 2002 (Wolf et al., 2017). These systems aim to help seniors maintain autonomy in their homes and improve well-being (Dohr et al., 2010). In the "Technology-supported motivation to maintain activity and quality of life" (TUMAL) study, a self-assessment measurement box was developed for seniors to track their physical abilities. The tests included the Timed Up and Go (TUG) test and the 5x Sit-to-Stand (SST) test, both assessing participants' mobility. The TUG involves standing, walking three meters, turning around, and sitting back down, while the SST involves standing and sitting five times consecutively (Fudickar et al., 2020). Interviews showed that participants wanted immediate access to their test results (Fudickar et al., 2022). However, providing real-time feedback requires significant personnel resources, which hinders independent and regular use. A solution is to deliver results via mobile devices in the form of a chatbot. Chatbots are now widely used in healthcare, marketing, and education (Adamopoulou and Moussiades, 2020), with customer support being a key application to reduce personnel costs (Adam et al., 2020). A recent review on the use of chatbots among older adults in healthcare concluded that there is a lack of options designed specifically for older adults. They find that adjacent studies are mainly focused on home monitoring and cognitive impairments. Furthermore, they did not identify any studies in this field that were conducted in Germany (Zhang et al., 2024). The chatbot

Philipp, W., Gölge, A., Hein, A. and Fudickar, S.
Usability Evaluation of a Chatbot for Fitness and Health Recommendations Among Seniors in Assisted Healthcare.
DOI: 10.5220/0013139800003911

In *Proceedings of the 18th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2025) - Volume 2: HEALTHINF*, pages 483-490
ISBN: 978-989-758-731-3; ISSN: 2184-4305

developed in this work will address this need by visualizing and delivering the physical activity and test results audiovisually.

1.2 Research Goals

The TUMAL study revealed a clear need for seniors to receive immediate feedback on their test results and the current technological advancements in the field of chatbots make them a suitable alternative for personal discussion of the results, from a technical standpoint. However, little is known about if seniors accept the presentation of assessments results via such chatbots and if they are suitable for usage. The core research question aims to clarify this research need: **Do seniors aged 75 and above accept a chatbot designed to display health data and provide personalized recommendations?** To answer this, two subquestions are formed:

1. How do seniors rate the usability and ease of use of the chatbot?
2. What is their attitude toward the recommendations provided by the chatbot?

2 METHODS

In order to address the research question, a mobile chatbot application specifically for seniors is implemented and is evaluated regarding the user acceptance and usability with representative participants of the age group.

2.1 Chatbot Application

The chatbot's main function is to visually display information about physical activity and measurement results from the health monitoring system, while also delivering audio-visual feedback based on the data. Additionally, it offers personalized recommendations aimed at improving both the user's fitness and overall health. It serves as a proof-of-concept, exploring whether seniors over 75 find this type of technology useful and could see themselves using it. The requirements for the chatbot have so far been vaguely formulated, or only the purpose has been derived from the results of the TUMAL study. To specify the functions, the requirements will be further defined to derive technical objectives. Accordingly, there are two specific requirements. The first is the verbal inquiry of the measurement box results/activity level and the audiovisual transmission of this information. The second involves the provision of action recommendations from the chatbot based on the measurement box results or the activity level. From these two requirements, application scenarios will be defined in the next sections.

2.1.1 Scenario 1: Querying Information

The first application scenario involves querying information regarding physical activity. The query can pertain to the measurement box results or the user's activity level. In response to the query, the chatbot displays a graphic that describes the user's performance. The chatbot verbally provides key information as an assessment. For self-evaluation and motivation, the user is also shown the average rating. Furthermore, the current performance is compared with past values. Additionally, there should be an option to access information about the TUG and SST tests along with reference values in an overlay window. The standard workflow of this use case starts with the user initiating a conversation with the chatbot. After launching the chatbot, the user can verbally communicate their desired query and will receive the described response.

2.1.2 Scenario 2: Making Recommendations

The second application scenario involves communicating recommendations to the user. At the start of the conversation, the chatbot first asks a question regarding the user's well-being. If the response is positive, the chatbot subsequently provides fitness recommendations intended to motivate the user to become more physically active. However, if the response is negative, the user is advised to consult a doctor. Therefore, these recommendations are referred to as health recommendations. The action recommendations are only communicated to the user in the event of a deterioration in the measurement box results or the activity level, or if the user indicates feeling unwell. The flow of this use case also begins with the user initiating a conversation with the chatbot. Before the user can start a query, the chatbot immediately asks about the user's well-being right after the conversation begins. Depending on the course of the conversation, as described above, the action recommendations are derived from the measurement values and communicated to the user.

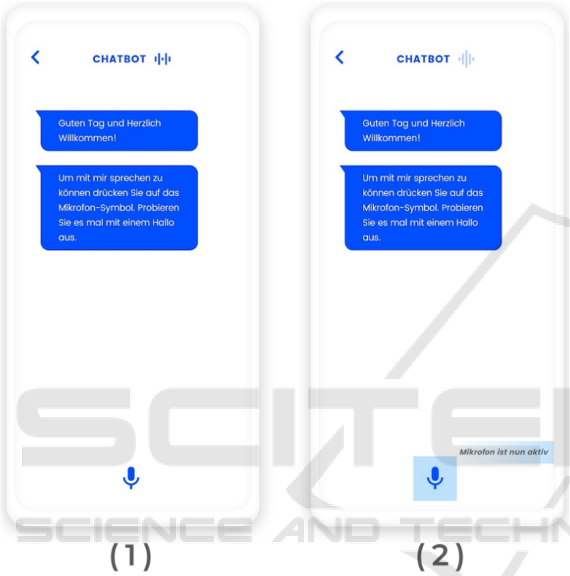
2.1.3 Design Guidelines for Seniors

The user group of seniors aged over 75 includes various factors that must be considered in the design of the user interface. To create a user-friendly and intuitive interface for this demographic, it is essential to incorporate design guidelines tailored for seniors.

Listing all aspects that were considered goes beyond the scope of this work. A sample of decisions that were made for the interface based on the user group were:

- Minimize the number of elements on the interface
- Use familiar elements from other apps
- Large elements and scalability
- No gesture controls

Figure 1 provides an example of how the design considerations were factored into the development of the application.



Guten Tag und Herzlich Willkommen!

Um mit mir sprechen zu können drücken Sie auf das Mikrofon-Symbol. Probieren Sie es mal mit einem Hallo aus.

Guten Tag und Herzlich Willkommen!

Um mit mir sprechen zu können drücken Sie auf das Mikrofon-Symbol. Probieren Sie es mal mit einem Hallo aus.

Mikrofon ist nun aktiv

Figure 1: Example of visual clarity implemented in the user interface. Left: The audiowave icon changes animation and colour when the chatbot is speaking. Right: The microphone icon lights up and displays the text ”microphone is active”.

2.2 Technology Stack

The chatbot was developed for Android due to its accessibility for developers and wide range of interfaces, making it suitable for mobile devices. Once integrated into a health app, the chatbot could access user activity data monitored in real-world scenarios. To meet technical requirements, offline data processing was prioritized. AIML (Artificial Intelligence Markup Language) was chosen to handle user inputs and generate outputs, as it is commonly used for chatbot development in research. For speech recognition (Speech-to-Text), the open-source API VOSK, based on the Kaldi toolkit, was selected due to its low memory usage and easy integration with Android. For speech synthesis (Text-to-Speech), Android’s built-in interface is used for simplicity.

2.3 Study Setup

With the chatbot application, a study is conducted with subjects from the target age group to evaluate its usability and assess how well the users accept the chatbot’s recommendations.

2.3.1 Study Group Selection

The recruitment process involved contacting a subgroup of 36 participants of the TUMAL study who agreed to be contacted for further studies by phone. 15 participants agreed to participate, of which 12 showed up for the study (see Table 1 for details on the makeup of the cohort). The participants were scheduled for appointments and invited to the university. Despite the smaller sample size, the methods used are expected to yield sufficient feedback to evaluate the chatbot’s usability.

2.3.2 Experimental Setup - Qualitative Phase

The study was conducted in individual sessions, with 40 minutes allocated per participant. The study required only a mobile device with the Android operating system, on which the application was installed. This device was provided to participants and positioned on a phone holder for convenient use. Both quantitative and qualitative metrics were established to measure during the study, with a focus on answering the research question and achieving the study’s objectives, particularly verifying the proof-of-concept. Participants were asked to evaluate not only the usability but also the general functionality and characteristics of the chatbot, such as voice clarity and speed, as well as satisfaction with the graphical display of information. Additionally, the study aimed to assess participants’ attitudes toward receiving recommendations from a chatbot. From a technical perspective, additional metrics were defined to evaluate the chatbot’s quality. The metrics are summarized in Table 2.

The first phase consisted of a practical user test, where participants had the opportunity to try out the chatbot. The usage involved having a conversation with the chatbot. No specific tasks were given to the participants, only the context was provided. The conversation consisted of two runs. In the first run, participants were given the freedom to choose between the two paths. In simple terms, this means that participants could freely respond to the chatbot’s question about how they were feeling with either ”Good” or ”Bad.” In the second run, participants were asked to choose the other option. The intention behind this was to show the participants the respective recom-

Table 1: Overview of the cohort that participated in the usability study.

Cohort	No. Participants	Avg. age	Min. age	Max. age	SD (age)
Men	8	82	76	90	4.03
Women	4	80	76	84	2.92
Total	12	81	76	90	3.77

Table 2: Metrics measured in the study.

Metrics	
Category	**Feature**
Qualitative	- Statements and questions from the Thinking Aloud method - Attitude towards recommendations
Quantitative	- Usability score according to the *System Usability Scale* - Usability results based on the *User Experience Questionnaire* - Satisfaction with the user interface - Satisfaction with the chatbot * Voice (speed, tone) * Speech recognition * Chatbot responsess - Satisfaction with the graphical representation of information - Number of matches between the user's speech and the system's understanding - Number of successful intent matchings - Error rate of intent matchings - Error rate of speech recognition

mendations provided by the chatbot. During the usage, participants were asked to follow the Thinking-Aloud method (JØRGENSEN, 1990). During this phase, a screen recording, including audio, was made for evaluation purposes. This allowed the tracking of user interactions and the documentation of statements according to the Thinking-Aloud method. Additionally, the chatbot application generated a log file in the background to record the conversation and store data related to speech recognition.

2.3.3 Experimental Setup - Quantitative Questionnaire Phase

Following the practical user test, the second phase involved a questionnaire survey. The questionnaire was provided to the participants in written form during the study. It began with personal information, such as name, age, gender, and a self-assessment of the participant's technical knowledge. The questionnaire is divided into three sections. The first section contains the User Experience Questionnaire (UEQ). The UEQ is a questionnaire designed to measure the usability of applications. It consists of a total of 26 questions that relate to six different metrics (Laugwitz et al., 2008). Each question presents two opposing attributes, and participants indicate their preference on seven item Likert scales. The next section covers the System Usability Scale (SUS). Participants respond to each statement using a five item Likert scale, ranging from ”Strongly disagree” to ”Strongly agree.” The statements consist of five positive and five negative ones. The total score, calculated using a predefined formula, ranges from 0 to 100 and reflects the user's perceived usability. A score of 68 or higher is considered indicative of good usability (Devy et al., 2017). The final section consists of six questions regarding the satisfaction with qualitative attributes, as listed in Table 2. Participants again had the option to express their satisfaction on a five-item Likert scale ranging from ”Not satisfied at all” to ”Very satisfied.” Additionally, participants were asked about their preference for a voice, with options including male, female, or no preference. Finally, participants were asked to assess how well they believed they managed the operation of the system.

2.3.4 Experimental Setup - Qualitative Interview Phase

At the end of the study, a short interview was conducted with the participants. The aim of this conversation was to understand the participants' attitudes towards the chatbot's health and fitness recommendations. This referred not to the content of the recommendations, but to the general acceptance of a technical recommendation system for health in the form of a chatbot. Additionally, participants' overall opinions about the chatbot were gathered.

3 RESULTS

The results of the conducted study are summarized in this section. During data analysis, the audio recordings from the practical user test were first transcribed. Evaluations and questions expressed during the Thinking-Aloud method were recorded in an Excel sheet for analysis. Based on this data, difficulties encountered during the use of the system were identified. The SUS score and UEQ result were calculated to quantitatively assess usability. Key metrics included the mean and standard deviation of both results. The data analysis focused on three key aspects. The first aspect was the measurement of usability, which was derived according to the methods presented in the previous section. The quantitative evaluations from the SUS score and UEQ result were supported by qualitative insights gathered through the Thinking-Aloud method. The second aspect was the participants' attitude towards the chatbot's recommendation feature. Interview transcripts were used for this evaluation. Finally, the third aspect focused on analyzing the speech recognition technology. The log files were compared with the transcriptions to determine the reliability of the speech recognition. It was also important to analyze how many words the participants used per input to understand their usage behavior.

3.1 Quantitative Evaluation

First, an overview of the qualitative results is presented in Table 3. The cohort rated their own technical knowledge on a five-item Likert scale ranging from ”Very low” to ”Very high,” with an average score of 4 (=Good). The participants' responses to the question ”Did you manage to use the system?” also resulted in an average score of 4.08 (=Good) on a five-point scale. The cohort's satisfaction averaged 4.18 (=Good). Of the twelve participants, four indicated that they would prefer a female voice. The remaining participants stated that they had no preference. Measured by the System Usability Scale (SUS), the usability resulted in an average score of 66.3. According to SUS, a score of 68 or higher indicates good usability. The score determined here is slightly below this threshold, which suggests poor usability. Figure 2 illustrates the results in a box plot. Of the twelve participants, six scored above the threshold of 68, while the remaining six scored below. A wide dispersion is noticeable. The standard deviation is 16.8. The results from the User Experience Questionnaire (UEQ), summarised in Table 4, are as follows: The outcome was slightly positive overall. The highest score was achieved in the attractiveness of the app, with a mean value of 1.36, while the lowest score was for efficiency, with a mean value of 1.04. The maximum standard deviation was observed for attractiveness and stimulation, both with a value of 1.48. The minimum standard deviation was recorded for efficiency and dependability, both with a value of 0.67. Except for efficiency (which was below average), all other metrics scored above average compared to other products within their benchmark distribution.

3.2 Qualitative Evaluation

3.2.1 Thinking Aloud Method

The following difficulties in operating the chatbot were determined using the Thinking-Aloud method. It should be noted that this methodology, when analyzed both quantitatively and qualitatively, did not yield significant results. The reason for this, according to observations, was the overwhelm experienced by many participants while using the chatbot. Consequently, most of the insights gained from this method consist of questions about the chatbot's usage, which were asked during the practical portion. The following list contains an excerpt of common problems the users experience and is ordered by frequency of occurrence. Each difficulty is accompanied by an example from the transcription of the audio recordings:

- **Microphone issues**: Participants either forgot or were unsure about pressing the microphone icon before speaking to the chatbot to activate it (*”Should I have pressed it again first?”*). This led some participants to initially think that the chatbot had not understood them (*”It didn't understand, did it?”*).
- **Touchscreen operation issues**: Some participants had trouble using the touchscreen (*”Pressing here isn't working so well.”*). Several clicks were not recognized, which led to further issues. For instance, when trying to activate the microphone, it wasn't clear to participants that their click hadn't been recognized, and they spoke into the microphone anyway. In this context, it's worth noting that visual indicators do appear when the microphone is activated.
- **Confusion between microphone and info buttons**: Despite clicking on the info button and an overlay window appearing, participants continued speaking as if the microphone was active (*”(Click on Info button) - What is my activity level?”*). Furthermore, participants were unsure how to interact with the information window (*”Where do I need to press here?”*). In one case, the information

Table 3: Overview of quantitative results. The participants are sorted by SUS score, descending.

#	Participant	Gender	Age	Tech. Knowledge	SUS	UX Rating	Avg. Satisfaction	Pref. Voice
1	P26	M	77	5	95.0	5	4	None
2	P38	M	83	4	87.5	5	4.5	Female
3	P36	F	84	1	77.5	4	4.5	Female
4	P4	M	90	5	75.0	4	4.5	None
5	P9	M	83	3	75.0	4	3.83	None
6	P8	F	76	4	70.0	5	4	None
7	P19	M	83	2	65.0	4	4	None
8	P31	M	80	1	60.0	3	3.83	None
9	P16	M	76	2	52.5	4	4	None
10	P20	F	79	2	50.0	4	4	Female
11	P40	M	82	4	47.5	3	3.5	Female
12	P34	M	80	2	40.0	3	3.83	None

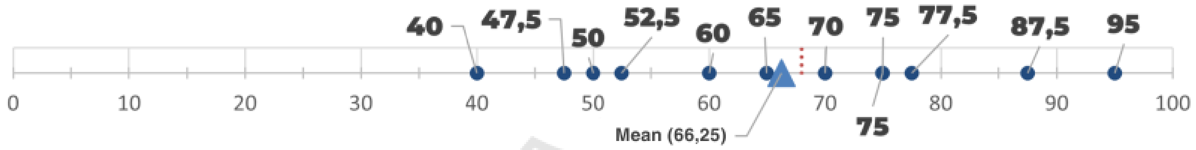


Figure 2: SUS scores plotted along a line. The cutoff point for good usability is marked in red.

Table 4: UEQ Results (Tabular). AA and BA denote above and below average results when compared to the benchmark, respectively.

Measure	Mean	SD	Benchmark
Attractiveness	1.36	1.48	AA
Perspicuity	1.33	0.90	AA
Efficiency	1.04	0.67	BA
Dependability	1.15	0.67	AA
Stimulation	1.24	1.48	AA
Novelty	1.08	1.38	AA

provided about the SST and TUG tests was interpreted by a participant as an instruction to perform the tests (*”What should I do now? Stand up and do the exercise...”*). The confusion was not limited to buttons. One participant, for example, mistook a speech bubble for a button or failed to realize that they had already clicked a button, and the subsequent speech bubble was displaying the result of that input (*”What should I press now (...) what should I press now?”*).

3.2.2 Interviews

The structured interview protocols provided insights into how participants viewed the fact that a chatbot gives health and fitness recommendations. The evaluation of the protocols led to the following findings:

- Eight participants found the chatbot’s fitness recommendations to motivate more physical activity as useful.
- Only five participants expressed a positive attitude toward the chatbot’s health recommendations and considered them unproblematic, noting that seniors regularly visit a doctor. However, these recommendations should be accepted with caution.
- Counterarguments included:
  - Not carrying a smartphone regularly.
  - A negative attitude toward technology in general.
  - Receiving recommendations based only on activity data was not viewed positively. The chatbot should collect more personal health data.
  - Regular visits to the doctor made the chatbot’s health recommendations seem unnecessary.
  - Health recommendations were seen as too extreme by some participants.
- There was consensus that the chatbot is a good idea for querying results after using the measurement box.
- Suggestions for improvement:
  - More dialog options.
  - The chatbot should provide information on why increased activity can positively impact participants and offer general health information.
  - The user interface (UI) should be better adapted for non-smartphone users using larger displays.

3.3 Evaluation of Speech Recognition

The following results emerged from the evaluation of the recordings and their comparison with the log files. Over the course of the practical part of the study, the twelve participants made a total of 127 inputs, comprising 694 recognized words. The average length of these inputs was five words. The Speech-to-Text API, VOSK, correctly recognized 75 out of the 127 inputs without errors (equaling 59.06% accuracy). A single incorrectly recognized word was counted as an error, with 97 out of the total 694 words being incorrectly recognized, resulting in a 13.98% error rate. A total of 531 words were recognized with a confidence level of 100%. The average confidence was 0.923, with a minimum value of 0.215. Despite low confidence, 65 words were correctly recognized.

4 DISCUSSION

4.1 Acceptance of Recommendations

Regarding the participants' attitudes toward the chatbot's recommendations, there was no clear consensus. Based on the findings, fitness recommendations aimed at motivating more physical activity were generally viewed more positively than health-related recommendations. This attitude could be attributed to a negative stance toward technology. Without a certain level of acceptance, it seems that trust in the technology is lacking, which makes the health-related recommendations from a chatbot seem irrelevant, especially for such a critical aspect of seniors' lives. Furthermore, many seniors mentioned that they are regularly under medical supervision and thus perceived these health recommendations as unnecessary from the start. One participant explained their negative stance based on observations of their social circle. They categorized seniors into two groups: those who visit the doctor regularly and those who avoid confronting their potentially poor health status, which is why they don't seek medical advice as often. This hesitation was also reflected in the interviews, where poor test results were sometimes taken personally. Participants emphasized that recommendations should be presented cautiously, as they could lead to panic among users.

4.2 Usability

Despite the low SUS score, the participants' individual evaluation reveals a differentiated picture, which was further highlighted through the interpretative approach presented. This is reinforced by the findings from the Thinking-Aloud method. It became evident that using the chatbot was associated with a high uncertainty for some participants, as difficulties during usage surfaced through frequent questions. Another example of uncertainty relates to the confusion between buttons. Despite the stark visual differences between two buttons, they were still confused. The results of the UEQ further underscore these usability difficulties. No significant correlation between the SUS score and other quantitative metric has been found. This is particularly interesting for the subjective metric of ”Technical Knowledge”, where participants where asked to rate their own technical knowledge. For instance, participant P40 rated the chatbot with a score of only 48, even though they assessed their technical knowledge as ”good”. Conversely, another outlier can be seen in Figure 2, challenging this assumption. Participant P36 rated their technical knowledge as ”low,” yet still awarded a SUS score of 78.

4.3 Speech Recognition

The speech synthesis created using the VOSK API proved to be a viable mobile solution that operates entirely offline. Despite the use of a small language corpus, an error rate of 14% was achieved. Considering that the participants were less tech-savvy compared to younger age groups, this error rate appears acceptable. However, when using AIML, which is responsible for output generation, the issue arises that even a single misrecognized word can cause the intent-matching to fail. Given the relatively long average length of inputs, the likelihood that a query results in no input matching is quite high. To mitigate this issue, the intents were cautiously designed based on the keyword method, which is why this problem did not appear in the results. The downside of this approach, however, is that conversations cannot be made more detailed. Therefore, a compromise must be found, using various matching methods to ensure a high likelihood of successful matching while also offering more diverse dialogue options.

5 CONCLUSIONS

It can be concluded that the general acceptance of the chatbot is evident. The interviews revealed that a majority of the participants considered the chatbot a good idea for checking results after using the measurement box. Based on the results, it was determined that the below-average usability rating may correlate with the technical proficiency of the participants. To avoid confusion with input methods, it appears necessary to limit users to one form of input. Currently, users can input commands either through voice or buttons, depending on the context. This switching between input methods caused confusion for some participants, which should be avoided. Supplementary features, such as displaying extra information regarding the TUG and SST tests, should be fully integrated into the chatbot. It was found that pop-up windows caused users to lose track of the interaction flow. Further usability improvements can be made according to the suggested enhancements. These include increasing the range of dialogue options, delving deeper into personal data queries for formulating recommendations, and supplementing the recommendations with explanations that justify them. Regarding the acceptance of recommendations, it would be better to limit the chatbot's advice to fitness-related suggestions. This might be achieved by considering the "mobility and endurance", "strength" and "balance", as main components to be considered in these assessments (Hellmers et al., 2017). Concerning the technology used, there is a need for improvements due to the demand for more dialogue options. The current AIML (Artificial Intelligence Markup Language) is error-prone due to its strict rules. As highlighted in the results, even a single error in the input can cause the intent-matching to fail. A potential solution would be to insert an additional module between the speech recognition and AIML systems. This module could function to improve the linguistic quality of the inputs. By addressing grammatical and spelling errors, this would reduce input errors and make the intent-matching more reliable.

ACKNOWLEDGEMENTS

This work was supported by the German Federal Ministry of Education and Research (BMBF) under grant agreement no. 16SV8958 and 1ZZ2007.

REFERENCES

Adam, M., Wessel, M., and Benlian, A. (2020). Ai-based chatbots in customer service and their effects on user compliance. *Electronic Markets*, 31(2):427–445.

Adamopoulou, E. and Moussiades, L. (2020). *An Overview of Chatbot Technology*, pages 373–383. Springer International Publishing.

Devy, N. P. I. R., Wibirama, S., and Santosa, P. I. (2017). Evaluating user experience of english learning interface using user experience questionnaire and system usability scale. In *2017 1st International Conference on Informatics and Computational Sciences (ICICoS)*, pages 101–106.

Dohr, A., Modre-Opsrian, R., Drobics, M., Hayn, D., and Schreier, G. (2010). The internet of things for ambient assisted living. In *2010 Seventh International Conference on Information Technology: New Generations*. IEEE.

Fudickar, S., Hellmers, S., Lau, S., Diekmann, R., Bauer, J. M., and Hein, A. (2020). Measurement system for unsupervised standardized assessment of timed "up & go" and five times sit to stand test in the community—a validity study. *Sensors*, 20(10):2824.

Fudickar, S., Pauls, A., Lau, S., Hellmers, S., Gebel, K., Diekmann, R., Bauer, J. M., Hein, A., and Koppelin, F. (2022). Measurement system for unsupervised standardized assessments of timed up and go test and 5 times chair rise test in community settings—a usability study. *Sensors*, 22(3):731.

Hellmers, S., Steen, E.-E., Dasenbrock, L., Heinks, A., Bauer, J. M., Fudickar, S., and Hein, A. (2017). Towards a minimized unsupervised technical assessment of physical performance in domestic environments. In *Proceedings of the 11th EAI PervasiveHealth Conference*, PervasiveHealth '17, page 207–216. ACM.

JØRGENSEN, A. H. (1990). Thinking-aloud in user interface design: a method promoting cognitive ergonomics. *Ergonomics*, 33(4):501–507.

Laugwitz, B., Held, T., and Schrepp, M. (2008). *Construction and Evaluation of a User Experience Questionnaire*, page 63–76. Springer Berlin Heidelberg.

Röcker, C. (2012). Smart medical services: A discussion of state-of-the-art approaches. *International Journal of Machine Learning and Computing*, pages 226–230.

Wolf, B., Scholze, C., and Friedrich, P. (2017). *Digitalisierung in der Pflege – Assistenzsysteme für Gesundheit und Generationen*, pages 113–135. Springer Fachmedien Wiesbaden.

Zhang, Q., Wong, A. K. C., and Bayuo, J. (2024). The role of chatbots in enhancing health care for older adults: A scoping review. *Journal of the American Medical Directors Association*, 25(9):105108.