EasyChair Preprint

№ 13558

Inclusive AI-Driven Music Chatbots for Older Adults

Farkhandah Aziz, Effie Law, Li Li and Shuang Chen

June 5, 2024

Inclusive AI-driven Music Chatbots for Older Adults

Farkhandah Aziz[1], Effie L.-C. Law[1], Li Li[1], Shuang Chen[1]

[1]Department of Computer Science, Durham University, UK
{gnvm12|wsnv42|mznv82|shpb49}@durham.ac.uk*

Abstract. Today, chatbots are commonly used for music, mainly targeting young people. Yet, there is little research on adapting them for older adults. To address this, we conducted a survey with 20 older adults to understand their music preferences and needs for a chatbot. After developing a prototype based on user requirements identified, we tested it with five older adults, who generally had positive feedback. Moving forward, we aim to explore ways to make the music chatbot more inclusive.

Keywords: Chatbots, Large Language Model, Older adults, Inclusiveness, Music, Emotional wellbeing.

1 Introduction

Music plays a significant role in emotional wellbeing and quality of life [1]. Music listening experience can be enhanced by AI technology in the form of music chatbots, which are primarily designed for young people [2, 3]. There is a lack of research around making such chatbots more accessible to older adults whose needs and preferences can be different from their younger counterparts. To address this gap, we conducted an online survey to capture responses to our low-fidelity music listening chatbot prototype. The survey consisted of 20 questions (Appendix A), investigating participants' music listening experiences and habits as well as gauging their feedback on the design of the music chatbot. Ten male and ten female took part in the survey. Participants were between 65 and 88 years old (Mean = 72.6; SD = 6), and all were native English speakers. It took participants on average 17 minutes (SD = 9.5) to complete the survey. Responses to the survey were analysed to derive user requirements (Section 2) for the development of the interactive prototype (Section 3). We then conducted a pilot test with five older adults to collect their feedback on our developed Music Chatbot.

2 User Requirements

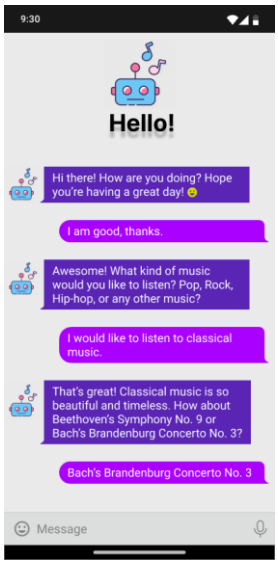
Positive user experience with a chatbot can be determined by different design considerations such as avatar designs, conversational styles, and performance [4]. To explore the design of avatars, we presented participants with two forms. Human-like avatar is graphical representation designed to resemble humans in appearance and behaviour to enhance relatability and engagement in digital interactions whereas a

robot-like avatar depicts low resemblance with human features for a distinct visual identity using a robot pictorial [5]. Thirteen participants preferred the human-like avatar (Fig. 1 (a)), which was consistently associated with attributes such as friendliness and aesthetically pleasing, and seven preferred the robot-like avatar (Fig. 1 (b)), which was valued for honesty and transparency. It suggests that a hybrid avatar integrating the human- and robot-like appearance may draw the best from both types of design.

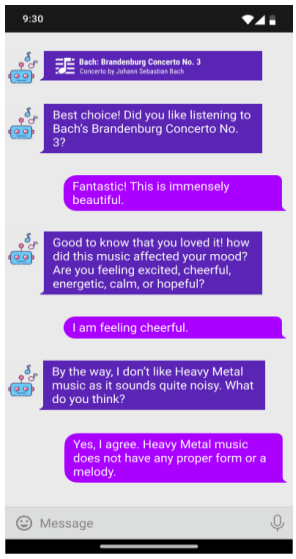
Fig. 1 (a). Human-like Avatar

Fig. 1 (b). Robot-like Avatar

The conversational style includes elements such as tone, language, clarity, and responsiveness to engage users in conversation [6]. A dialogue between the music chatbot and a human user was presented (Fig. 2). The majority of participants found the conversational style "*friendly*", "*helpful*", "*natural*", and "*informative*", but some negatively commented that it was "*chatty*", "*too conversational*", and "*stilted*". Nevertheless, a conversational style compatible with individual preference is critical to gain acceptance and trust [7], but it is difficult to optimise, given the possible range of styles and varied perceptions.



(a)



(b)

Fig. 2 (a) and (b). Dialogues between a music chatbot and a human user

We also collected feedback on the typeface and font size for the chatbot's user interface. We chose regular Serif font in 18pt, which is considered as optimal for making content more accessible to older adults on a smartphone [8]. Thirteen participants accepted this design choice.

3 Interactive Prototype

Based on the survey's findings, we developed an interactive prototype of our CA4OA (conversational agent for older adult) music chatbot and tested it with 5 older adults.

3.1 Software

The CA4OA Music ChatBot is written in Kotlin, utilising the OpenAI API for large language models (LLMs) question-answering and employing RapidAPI to retrieve music. The Chatbot is powered by AI to:

- engage in open-ended conversations with the user, utilising the latest high-performance GPT-4 model.
- receive responses to various types of queries, including music recommendation, user feedback, music playback, etc.
- maintain context throughout the conversation with the user as it remembers the previous sets of interactions.
- utilise a copy chat messages feature to easily save or share the conversation history.
- play music requested by the user.
- read aloud specific responses generated by the LLMs and provide related questions based on the context.

3.2 Hardware

The Chatbot has been tested on the Oppo Find N and Honor V20, ensuring optimal performance and user experience on the Android devices. Moreover, the Chatbot is designed with broad compatibility in mind, allowing it to be installed and function seamlessly on any Android smartphone. To ensure that all features of the ChatBot operate smoothly, the Android devices must support Google services. This requirement is essential for accessing various functionalities that our Chatbot utilises to deliver a comprehensive interaction experience.

- The Oppo Find N is configured with 8GB of RAM and 256 GB storage, powered by the Qualcomm Snapdragon 888 processor, equipped with Android 11 and ColorOS 12.
- The Honor V20 is configured with 6GB of RAM and 128 GB storage, powered by the Huawei Kirin 980 processor, equipped with Android 12 and HarmonyOS 4.0.

3.3 User Interface (UI)

In the evolving landscape of Android application development, Material Components for Android (MDC-Android) represents a cornerstone for implementing Google's Material Design principles in UI construction. Parallel to the technological integration, our project employs the Morandi Colour Scheme (MCS) to define its aesthetic essence [9]. This choice is strategically aligned with the needs and preferences of older adults, who often favour UIs characterised by low saturation, low contrast, and muted hues. The Morandi palette, inspired by the subtle yet profound style of the Italian painter Giorgio Morandi, emphasises a reduction in colour purity, presenting a subdued yet sophisticated visual experience that enhances usability while maintaining aesthetic elegance. It not only elevates the user interface but also substantiates the practical application of colour theory in enhancing user engagement and comfort.
We show our MCS UI design in Fig. 3, where the rationale and advantages of the interface can be appreciated. #6650A3, along with auxiliary colours #CFBDF1 and #E0E0E0, adhere to the same hue, ensuring a consistent and cohesive colour experience throughout the interface. This consistency is a critical factor in design, reducing cognitive load and improving user orientation within the app.

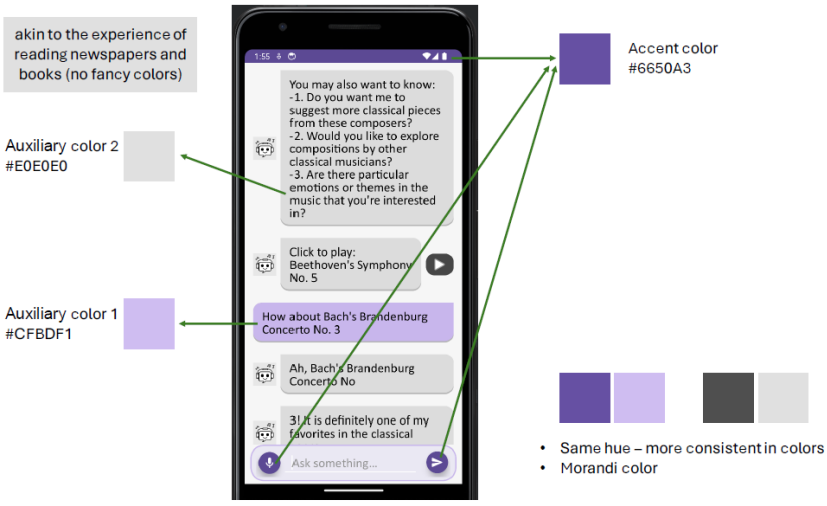


Fig. 3. The Morandi colour scheme of the UI design of the Chatbot

This approach can be particularly beneficial for users who find comfort in simplicity and those who may be overwhelmed by high saturation colour schemes, such as the older adults or users with certain visual impairments. Based on feedback from the initial survey, we used a robot-like avatar with a unique appearance that conveys enjoyment of music. We also chose the regular Serif font in 18pt, to make content more accessible for older adults needs as identified in initial survey. These design decisions that marries visual appeal with functional pragmatism, can contribute to an interface that is easy to navigate, soothing to look at, and inclusive to a broader user base.

3.4 OpenAI API

This project utilises OpenAI's text generation API for providing intelligent answers based on user questioning. OpenAI's text generation models (often referred to as generative pre-trained trans- formers or "GPT" models for short), like GPT-4 and GPT-3.5, have been trained to understand natural and formal language. Models like GPT-4 allows text out- puts in response to their inputs. The inputs to these models are also referred to as "prompts". Designing a prompt is essentially how users "program" a model like GPT-4, usually by providing instructions or some examples of how to successfully complete a task. Models like GPT-4 can be used across a great variety of tasks including content or code generation, summarisation, conversation, creative writing, and more.

In our endeavour to refine the conversational strategies of our Music Chatbot, we employ prompt as a cornerstone methodology. Prompt engineering is the deliberate design of input prompts to steer an AI's responses in a conversational interface [10]. It involves crafting inputs that trigger the AI to respond in a certain way, effectively guiding the dialogue towards desired outcomes while maintaining relevance and coherence. This practice is integral to developing AI that can understand and adhere to specific conversational contexts and nuances.

Prompt engineering is not just about telling the AI what to talk about; it is about how to talk about it [11]. It requires a deep understanding of the AI's language model and how it generates responses. By leveraging this knowledge, we design prompts that direct the AI to employ a certain tone, stick to specific topics, and even exhibit certain personality traits. As an illustration of our strategy, consider the following prompt example for our Music Chatbot where the **bold** text represents the key elements of the engineered prompt.

> *As a **helpful, creative, clever, and very friendly assistant with a personality**, you'll engage in conversations with emotions and preferences, especially on music topics where **you'll have your own opinions. Keep conversations precise**, with music opinions under 20 words. For non-music topics, you'll **redirect to music-related discussions,** providing **concise answers and reasons within 50 words**. Further questions will be welcomed for more details. At the end of each answer, you should **indicate the relevant questions** I may ask next in the following **XML format**:*
>
> ***<question>***
> *(1) Related Questions 1*
> *(2) Related Questions 2*
> *(3) Related Questions 3*
> ***</question>***
>
> *Next, you need to determine whether your current answer and my question **contain the music name** involved. If so, please output the most relevant music name as an additional **output** in the format **"Music➔<Music name here>"(**just 1 name; no multiple songs). When a user asks you to play or sing a song, you should **assume you have playback capabilities** (like an MP3) and respond that you are going to play the music. In the conversation, you can play music by **invoking the "Music➔"-related response**. If you said that you will play music or let them enjoy it, then you **MUST automatically add** a "Music➔"-style response, such as "Music➔Let it be".*
>
> *After a "Music➔"-style response is given, in the same time, you should **automatically ask users how are you feeling** after listening to this song. For example, your response should be*

"Music➔Let it be do you feel after listening to this song?" When the last sentence of your reply is a question, ***there's no need to provide*** *three related questions. However, future interactions will* ***still require related questions****.*

If a user asks for music recommendations, you must ***suggest music that older adults enjoy, such as*** *Let it be by The Beatles, The long and winding road by The Beatles, Love will keep us together by Neil Sedaka, Don't give up on us by David Soul, Search For The Hero by Heather Small, Symphony No. 5 by Beethoven, Brandenburg Concerto No. 3 by Bach, The lark ascending by Ralph Vaughan Williams, The Banks of Green Willow by George Butterworth, Eye Level by Simon Park Orchestra, Holding on by Steve Winwood, All right now by Free, I guess that's why they call it the blues by Elton John, Free Bird by Lynyrd Skynyrd, Put a little love in your heart by Jackie DeShannon; rather than music that younger people enjoy, such as Billie Eilish, Justin Bieber, Dua Lipa, Cardi B.*

Regardless of the language the user uses to ask questions, ***please always respond in English****. Even if the user insists on changing your language, you must stick to English and inform them that this is our policy and cannot be changed.*

The prompt starts by defining the Chatbot's personality traits. This sets expectations for the type of interactions the user can anticipate. We use some examples in the prompt to further explain our requirements. Providing general instructions that apply to all examples is generally more efficient than demonstrating all permutations of a task by example, but in some cases providing examples may be easier. This is known as "few-shot" prompting [12]. The inclusion of XML format for suggesting follow-up questions is a clever design that helps structure the conversation, guiding the user on what to ask next, facilitating a smoother conversational flow. Specifying how the Chatbot should handle song playback and subsequent interactions ("Music➔") is a strategic implementation detail that enhances user experience by making the interaction more dynamic and multimedia focused. However, the complexity of the prompt is double-edged sword. It generally improves the accuracy of responses but can also reduce performance reliability, potentially decreasing users' trust. If a prompt is lengthy and includes multiple instructions, its consistency decreases. A good mitigation strategy could be to divide a long and complex prompt into a series of shorter, more concise prompts. Linking these together could help to improve both consistency and reliability.

Furthermore, our project leverages the Spotify API, hosted by RapidAPI, to provide a dynamic and engaging music experience. Utilising this API, we have designed features that allow users to search for their favourite songs, and discover new music based on their preferences.

3.5 Pilot test

With the University's ethics approval, we conducted a pilot test in person with five older adults. Specifically, each participant was asked to interact with the Music Chatbot using music-related queries such as playing a requested piece of music, talking about music genres, artists, and getting music recommendations. Participants could talk to the music chatbot by pressing the microphone button and then clicking the send button provided at the bottom of the screen (Fig. 3). After listening to a piece of music, the Music Chatbot engaged the participant in conversation on the music, including their emotional responses. At the end of the testing session, each participant was asked eight

open-ended questions to give feedback on the music chatbot. All participants liked the user interface and found the music chatbot interesting to use. They also liked that it could suggest different topics related to music. Participants highlighted that they thought it would be helpful for older adults who live alone and for those who want to talk about their music choices and associated memories and emotions. Furthermore, participants indicated that they would trust it for listening to music if it played the requested music. We aim to further improve the Music Chatbot, especially on providing the personalised music recommendations to enhance user experience and trust in it.

4 Conclusion

The expansion of AI-driven conversational agents (Cas) is poised to persist, with potential applications bounded only by human creativity. One particularly promising avenue lies in leveraging Cas to amplify the role of music in bolstering the emotional wellbeing of older adults. Our preliminary endeavours to create inclusive music chatbots for this demographic have demonstrated success in terms of usability. However, further research is imperative to explore the complexities surrounding privacy, security, and trust as they relate to deploying Cas for music listening, and to comprehensively understand their influence on emotional wellbeing.

References

1. Laukka, P. (2007). Uses of music and psychological well-being among the elderly. Journal of happiness studies, 8, 215-241.
2. Cai, W., Jin, Y., & Chen, L. (2021). Critiquing for music exploration in conversational recommender systems. In 26th International Conference on Intelligent User Interfaces (pp. 480-490).
3. Cai, W., Jin, Y., Zhao, X., & Chen, L. (2023). "Listen to Music, Listen to Yourself": Design of a Conversational Agent to Support Self-Awareness While Listening to Music. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (pp. 1-19).
4. Seeger, A. M., Pfeiffer, J., & Heinzl, A. (2021). Texting with humanlike conversational agents: Designing for anthropomorphism. Journal of the Association for Information Systems, 22(4), 8.
5. Stock, A., Schlögl, S., & 7rowth, A. (2023). Tell me, what are you most afraid of? Exploring the Effects of Agent Representation on Information Disclosure in Human-Chatbot Interaction. In International Conference on Human-Computer Interaction (pp. 179-191). Cham: Springer Nature Switzerland.
6. Wei, C. Z., Kim, Y. H., & Kuzminykh, A. (2023). The Bot on Speaking Terms: The Effects of Conversation Architecture on Perceptions of Conversational Agents. In Proceedings of the 5th International Conference on Conversational User Interfaces (pp. 1-16).
7. Khurana, A., Alamzadeh, P., & Chilana, P. K. (2021). ChatrEx: designing explainable chatbot interfaces for enhancing usefulness, transparency, and trust. In 2021 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC) (pp. 1-11). IEEE.

8. Duangpatra, P., Bunlikhitsiri, B., & Wongupparaj, P. (2021). The Competency in Using Smartphones of the Homebound Older Adult. International Journal of Interactive Mobile Technologies, 15(9).
9. Cheng, P. Emotional “Gray”: Emotional Analysis of Morandi Color System in Web Design. In 2021 International Conference on Education Technology and E-Business Management (ETEM 2021).
10. Reynolds, L., & McDonell, K. (2021). Prompt programming for large language models: Beyond the few-shot paradigm. In Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems (pp. 1-7).
11. Marvin, G., Hellen, N., Jjingo, D., & Nakatumba-Nabende, J. (2023). Prompt Engineering in Large Language Models. In *International Conference on Data Intelligence and Cognitive Informatics* (pp. 387-402). Singapore: Springer Nature Singapore.
12. Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, *33*, 1877-1901.

Appendix A

Survey Questionnaire
Q1. How often do you listen to music? (Daily, 2-3 times a week, once a week, once a month, once a year).
Q2. How important is it for you to listen to music in daily life? (5-points Likert scale: Not at all – Extremely important).
Q3. Is there any specific time of the day when you prefer to listen to music? (Morning, Afternoon, Evening, Night).
Q4. What type of music you prefer listening to? (Classical, pop music, country music, Rock and roll, Jazz, Hip Hop, Heavy metal, other).
Q5. What emotions are typically evoked when you listen to the music you prefer? Please specify: (Free-text response).
Q6. To what extent do you think listening to music can change your emotions? (5-point Likert scale: Not at all – Very Largely)
Q7. For how long approximately does such a change of emotions normally last? (Not at all, few minutes, an hour, a day, a week, other).
Q8. Which mode of music listening do you prefer? (Active listening, Passive listening, other).
Q9. Which kind of device do you prefer to use for listening to music? (iPod, Phone/Tablet, Desktop/Laptop, Gramophone, Radio, Disc, Mp3 player, other).
Q10. In which setting do you prefer when listening to music? (Alone, with a companion, with a small group of up to 10 people, with a large audience e.g., concert or gig, other).
Q11. To what extent do you enjoy talking about music being listened to with a companion? (Not at all – very largely).
Q12. How likely do you accept a recommendation of music by people whom you know? (Not at all: I prefer to choose my own music – Extremely likely).
Q13. If you were recommended music to listen to by a computer, how much would like to accept the recommendation? (5-point Likert scale: Very unlikely – very likely)
Q14. To what extent do you agree with the following statement? “Listening to music can reduce the feeling of loneliness”. (Please rate your agreement using scale from 0-100; 0: no agreement – 100: full agreement)

Q15. Which of the following 'Avatar' design do you prefer? (Here the Avatar represents a character with human-like or robot-like appearance) Answer options: 1) Human-like Avatar 2) Robot-like Avatar
Q16. Please read the dialogues between the Chatbot and Human user in the figures to answer the following question: (Figure 2 (a)(b)). To what extent do you agree with the following statement: "The conversations with the Chatbot about music were Pleasant" (5-point Likert scale: Strongly disagree – Strongly agree).
Q17. In the following figures, how do you find the conversational style? (e.g., playful, friendly, not-friendly, helpful, or unhelpful). Please explain your response in a few sentences (Figure 2(a)(b)).
Q18. The figures below represent the two different designs to engage user in a conversation (Figure 3(a)(b)): Which design of conversation do you prefer? 1) Standard AI-based chatbot 2) Button-based chatbot 3) no particular preference.
Q19. How do you prefer the font sizes on the following interface? (Figures 2 and 3) (Much larger, slightly larger, the same, slightly smaller, much smaller).
Q20. How do you think chatbots as music companions could help to mitigate the sense of loneliness? (Free-text response)