Designing Personalities for Mental Health Conversational Agents

Joonas Moilanen[1], Aku Visuri[1], Elina Kuosmanen[1], Andy Alorwu[1] and Simo Hosio[1]

[1]Center for Ubiquitous Computing, University of Oulu, P.O. Box 4500, FI 90014, Finland

Abstract

In recent years, the use of conversational agents (CA) has been increasing at a rapid pace. Efforts have been made to leverage CAs for tackling mental health challenges. Our goal is to improve mental well-being by enabling self-help ideas through chat-based CAs. To enhance the effectiveness and user reception of CAs, we designed different conversational personalities with low and high variants of extroversion and conscientiousness. We used various language cues and example conversation scripts as the basis of our design process. This paper presents the design and validation process of such CA personalities. Our results indicate that the final personality characteristics presented in the scripts are recognizable in the text-based CA interactions and thus enable future research on human behavior with such agents.

Keywords

Conversational agent, chatbot, mental health, personality, conversational design

1. Introduction

Research on the use and benefits of Conversational Agents (CAs) for health and well-being has been steadily increasing [1]. Many CAs are also used in communication with customers on commercial platforms and are increasingly utilized in official government websites and general healthcare [2]. In this paper, we focus on chat-based CAs, so-called chatbots, in the context of mental health self-help. CAs have been used in mental health self-help applications, yielding promising results in promoting self-help and reducing stress [3, 4, 5].

If the CA appears too human, it can significantly lower the user's trust in the CA and make it seem uncanny, an effect first suggested by Mori [6] and later researched and discussed in several other papers [7, 8, 9]. Lately, researchers have focused on how different personality types perform and how using different personalities or even altering it depending on the user can make the CAs perform better by increasing the user's trust and acceptability towards them [10, 11, 12, 13, 14, 15]. Focusing on the personality design can also help make them more easily approachable and likable, thus increasing their effectiveness [16, 17]. Matching the CA and user personalities can have a positive effect on the user experience [18] and many users wish for CAs to show some kind of personification, expect them to show emotion and more attention to the user [19] and treat them like a human being when engaging in a conversation [20].

Beyond the work presented in this paper, we are interested in researching how CAs can be used for mental health self-help. We aim to study how different CA personalities affect the users' opinion on the CA and how the user's personality affects their perception of the CA personality. In a mental health context, users can have strong reactions in social interaction due to, e.g., past traumas being triggered from specific types of behaviors, which should be avoided at all costs. Thus, ensuring no unwanted reactions ensue when interacting with a CA is critical. This paper presents the process of designing five different CA personalities using conscientiousness and extroversion from The Big Five personality traits [21]. We also present the results of our iterative validation process of the created personality types. Using this process, we created five sufficiently distinct personality types for our CAs, with personality features identifiable from each other by crowdworkers. In future work, we will evaluate how each of the designed personalities matches with end-user personalities and how well-suited each designed personality is for offering mental health self-help ideas through a chatbot.

2. Related Work

Using CAs to provide mental health support, self-help, and digital counseling services has been researched in various studies [22, 23, 24, 25]. CAs can help mental health counseling services to be more accessible. CAs can help users engage in conversation and improve well-being in non-clinical use [3]. Self-help CAs provide the users with instant and interactive feedback as opposed to traditional websites and can be beneficial, especially for

Joint Proceedings of the ACM IUI Workshops 2022, March 2022, Helsinki, Finland

joonas.moilanen@oulu.fi (J. Moilanen); aku.visuri@oulu.fi (A. Visuri); elina.kuosmanen@oulu.fi (E. Kuosmanen); andy.alorwu@oulu.fi (A. Alorwu); simo.hosio@oulu.fi (S. Hosio)
0000-0002-5750-4918 (J. Moilanen); 0000-0001-7127-4031 (A. Visuri); 0000-0001-6725-8848 (E. Kuosmanen); 0000-0002-7210-2896 (A. Alorwu); 0000-0002-9609-0965 (S. Hosio)

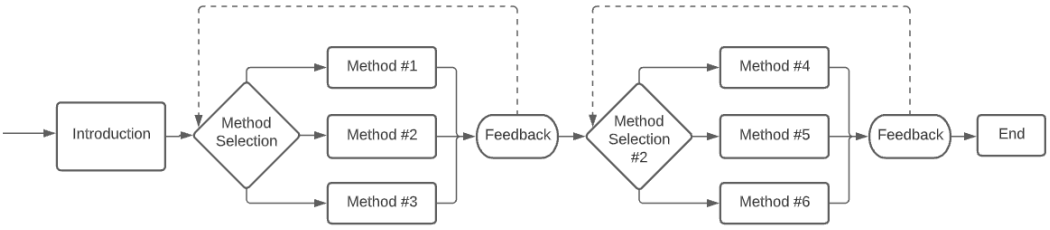

Figure 1: The conversation flow used for all personality types. Before feedback on the provided self-help method(s), the user can revisit any previously unselected methods.

individuals who prefer to seek their information online [22].

Research has shown that users prefer to engage with CAs close to their personalities for information-seeking [26] and in assistance and therapy [27]. Having the CA express multiple personalities and have emotion selection improves the user engagement significantly [12]. In particular, the effect of high extroversion has been shown to enhance user satisfaction [10].

3. Conversational Agent Design

We created five CAs with differing personalities that offer self-help method recommendations for mental health. To appropriately compare one personality to another, we wanted to minimize effects that can be considered external. One of the factors identified was the amount of information provided by the CA. To mitigate this, we designed a simple interactive conversation structure that provides the user with mental self-help methods, which all CA personalities follow. The main conversation structure is shown in Figure 1.

After the initial introduction, the user can select their current mood from three options - *Not so good*, *I'm okay*, and *I feel great*. The CA reacts to the chosen option accordingly and gives three mental health topics, including stress, anxiety, and low mood, that it gives self-help methods for. This step can be repeated as many times as the user wants, after which the user is asked for an open text entry about what they thought of the methods given. The CA responds to this message accordingly. The second part of the conversation follows a similar pattern, with the CA giving information on literature, audio, and website sources.

3.1. Designing Conversational Agent Personalities

An individual's personality is a combination of multiple attributes and cues, ranging from behavioral to emotional and mental characteristics. Personality can manifest as personality traits, of which the typical dimensions are the so-called Big Five personality dimensions [21]:

- Extroversion vs. Introversion (sociable, assertive, playful vs. aloof, reserved, shy)
- Emotional stability vs. Neuroticism (calm, unemotional vs. insecure, anxious)
- Agreeableness vs. Disagreeable (friendly, cooperative vs. antagonistic, faultfinding)
- Conscientiousness vs. Unconscientious (self-disciplined, organized vs. inefficient, careless)
- Openness to experience (intellectual, insightful vs. shallow, unimaginative)

The Big Five traits have been leveraged as analysis tools, for example, for academic success [28] and in mental health contexts [29]. Language use is frequently utilized as a method of identifying Big Five personality traits [30, 31]. However, some of its terms like neuroticism and language associated with such behavior can vary strongly between individuals. We wanted to leverage the Big Five personality traits in the design of our CAs and considered the following before beginning the design process.

Firstly, the measurements or factors used to determine Big Five traits can overlap, e.g., factors perceived as high on the extroversion-introversion dimension can also exist on the agreeable-disagreeable dimension. Thus, if a CA personality type A exhibits features of high extroversion, the same features can be perceived as, e.g., high agreeableness. This persuaded us to consider the CA personalities as single dimensions instead of a combination of all Big Five or a subset of them - i.e., hypothetical CA personality "A" has *high extroversion*, and its personality

is measured only on the extroversion scale. This division of the Big Five into dimensions is a standardized method of personality trait analysis [14].

Secondly, not all the dimensions are necessarily suitable or relevant for a CA designed to offer self-help guidance. E.g., openness to experience as a trait of the CA plays a minor role in how i) the CA delivers information and ii) how this information is received. Thus, we wanted to select a subset of the dimensions to analyze and decided on the *extroversion* and *conscientiousness* traits. Reaction to social assertiveness (or lack of it) is shown to vary according to an individual [32] thus how assertive (or shy) the CA personality is can affect how it is perceived. Conscientiousness can also be perceived differently by individuals [33], as high conscientiousness can have a negative influence on individuals with depression [34] but is generally considered a positive trait. In earlier CA-related personality studies, we found 10/11 articles to study extroversion-related traits, 5/11 to investigate emotional stability-neuroticism - which we would find somewhat unfitting for our purposes, especially on the neurotic end, and 4/11 studied conscientiousness. Both of the selected dimensions are evaluated on the high-low scale, so the selected personality types are High and Low Extroversion (**High E** and **Low E**) and High and Low Conscientiousness (**High C** and **Low C**), as well as the Neutral option.

These CAs follow the same conversational structure and differ only in their personalities, using different phrases and words to deliver the same information. In addition to these, we created a CA with a neutral personality that aims to be neutral in both conscientiousness and extroversion. We started by making the neutral conversation script, then created the other personalities by using various language cues, some of which can be seen in Table 3, and example conversation scripts of different personalities from related work [14, 35]. To help users differentiate between the CAs, we gave each CA an avatar and a color theme.

3.2. Deployment

We deployed the chatbots using BotStar chatbot engine[1] and Dialogflow ES[2]. BotStar was used to create the main conversational structure and host the CA. In contrast, Dialogflow ES was used for two specific messages that determine if the user is satisfied with the recommendations given and has the CA respond to their message accordingly.

To keep the design simple and to prevent the user from selecting topics the CA is not able to talk about, the primary interaction between the user and the CA is through multiple-choice buttons (method selection) or by natural language processing methods provided by Dialogflow ES. The latter allows the CA to react in a few different ways to the participant's response to "What did you think of the presented methods?" during the feedback, for example, which is responded to using text input.

[1]https://botstar.com/
[2]https://cloud.google.com/dialogflow

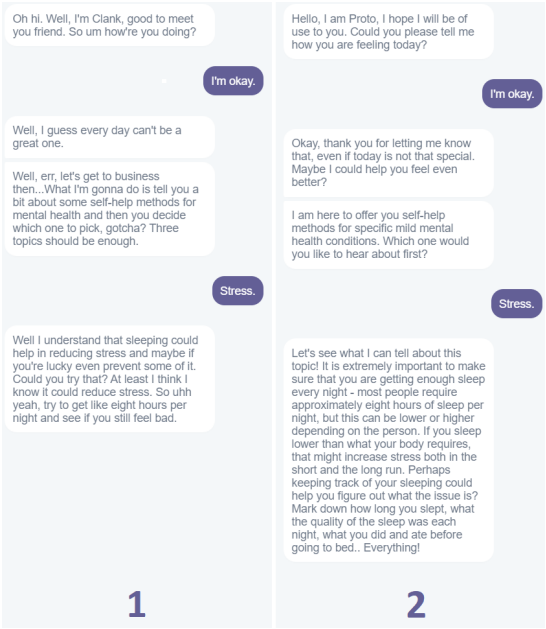

Figure 2: In the questionnaire, the participants were presented with two conversations and were required to grade the language cues and overall conscientiousness/extroversion on a 7-point Likert scale. In this example, 1 = Low conscientiousness 1 and 2 = High conscientiousness.

4. Personality Validation

The conversation scripts of the five different CAs were created by the main author of the paper and then evaluated by two co-authors. We then designed a questionnaire to validate the personality design and used the online crowdsourcing platform Prolific [36] to recruit participants. Depending on the results we gain from the personality validation, we can change the conversation scripts accordingly and re-validate. To locate potential issues with precision, we select three conversation parts from each personality and three individual messages from the remaining script. This helped us see which specific section of the script needed to be re-written instead of re-doing the whole script. For this purpose, the participants do not directly use the CA, and the entirety of the validation is done by presenting the users with

images of conversations as is presented in Figure 2 and singular messages directly on the survey platform.

4.1. Questionnaire Design

For the validation process, we designed a three-stage questionnaire. To ensure each personality is perceived as intended, we validated both Conscientiousness and Extroversion dimensions separately - High C vs. Low C and High E vs. Low E. To ensure the participants are familiar with the two personality types, we briefly described the personalities and presented the participants with example messages that portray each personality type. In sections 1 and 2, we presented the same conversation side-by-side, as shown in Figure 2. We then used a single-choice selection to evaluate three language cues used in part 1 and part 2 (*"Which set of the white messages is more Impulsive?"*). White messages refer to the text in the conversation outputted by the CA. The range of language cues varied from more obscure terms like "Vague," "Impulsive," "High verbal output" to more understandable terms like "Formal" or "Using shorter words." The language cues we used were derived from the work of Mairesse et al. [14] and are based on several studies researching different Big Five traits [37, 38, 39, 40, 41, 42, 43, 44, 45]. The full list of language cues are presented in Table 3. Then, we evaluate the conscientiousness or extroversion of the script using a single 7-point Likert-style Item (*"How conscientious do you think set 1 was?"*). This method was used for three conversation sets for both Conscientiousness (Questionnaire section 1) and Extroversion (section 2). In section 3, individual messages were presented, and the participant was instructed to select a matching personality type (Neutral, High C, Low C, High E, or Low E).

4.2. Participant Recruitment

Participants were recruited on the Prolific crowdsourcing platform. Prolific helps reach out to individuals from different backgrounds and with different native languages, which can be an important factor when evaluating CA conversations made in English. We used the platform's pre-screening tool to limit the participants to those with an approval rate of at least 95%, at least 50 previous submissions, and excluded the participants of the first iteration from the second.

5. Results

Creating exact and universally perceived types of personalities can be difficult. Interpreting and rating more obscure language cues such as "Impulsive" and "Vague" can vary for each individual. Rating the overall conscientiousness or extroversion should prove to be more accurate. As language cues between personalities can overlap, which can lead to, for example, Low E being misinterpreted as High C, section three of the questionnaire could prove difficult to get accurate results in.

40 participants were recruited, 20 for each iteration. 3 participants were replaced for timing out or failing to answer the control question. The average age of participants was 27.23 years (SD = 7.49). 24 were female, 16 male. 31 of the participants came from Europe, with Portugal and The United Kingdom being the most presented countries, with 7 participants for each. The remaining participants came from North America (4), South America (2), Africa (2), and Asia (1). English was the first language for 10 participants. The participants were paid 6.70USD/hr for the first iteration and 9.24USD/hr for the second iteration, with an average response time of 12min 11s and 10min 52s, respectively.

We now introduce the results of the first iteration for each section, the changes the responses elicited us to conduct, and finally, the results after the second iteration round.

5.1. First Iteration

The first two sections of the validation questionnaire rated the language cues and the overall conscientiousness/extroversion of three sets of conversations. To validate the language cues, we used the simple majority voting system, i.e., results where the correct answer is the most selected option are deemed sufficient. Only the language cue *"dissatisfaction"* for Extroversion set 3 resulted under this threshold, with 7/20 participants unsure and 6/20 correct.

For the Likert-style items, we visually inspected the distribution of answers, presented in Figure 3 and Figure 4. We then analyzed the differences between Low and High personality variants using the Wilcoxon Rank Sum test, and all the variants were significantly different on the .005 confidence level, full results presented in Table 2.

In section three of the questionnaire, we asked participants to validate individual messages of the conversation and select the personality type from all of the five possible options, including 'Neutral / I'm not sure.' Each personality has three messages to be evaluated, for a total of twelve messages. The results are presented in Table 1. Again, we used a simple majority to validate the results. Low E was most often selected as High C (21/60) and as Low E only 13/60 times. Low C was mistaken for High C 17/60 times and correctly selected 21/60 times.

5.1.1. Modifications

After the first iteration, Low C and High C were not as distinct from each other as we had hoped. In the Low

Table 1
Validation of the individual conversation messages. Each of the 20 participants responded to 12 multiple choice questions where each of the four options was correct three times (3x20=60). NA indicates the "I don't know / I am not sure" answer. Grey indicates the 'correct' answer.

	Iteration #1					Iteration #2				
	Low C	**High C**	**Low E**	**High E**	**NA**	**Low C**	**High C**	**Low E**	**High E**	**NA**
Low Conscientiousness	21	17	6	12	4	31	5	10	13	1
High Conscientiousness	3	41	5	9	2	2	46	0	9	9
Low Extroversion	9	21	13	6	11	13	17	16	8	6
High Extroversion	10	9	7	30	4	9	7	8	31	5

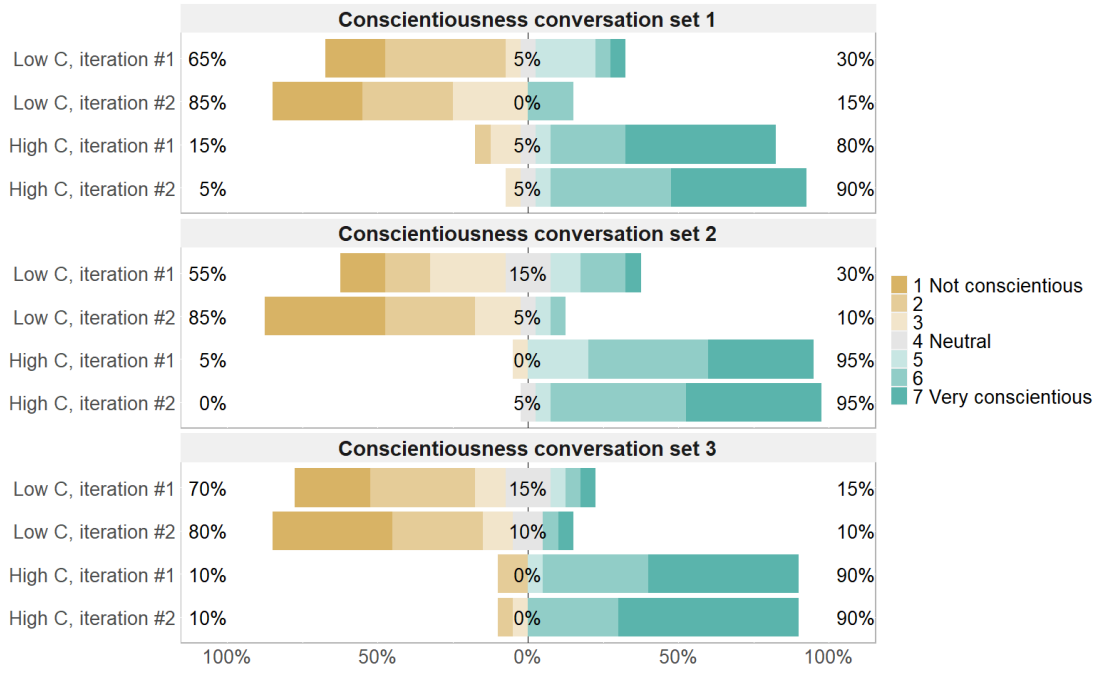

Figure 3: Results for the conscientiousness of the conversation sets in the first section of the validation for both iterations. The figure shows the proportion of the participants rating the conversation set neutral (4) in the center, and low (1-3) and high (4-7) conscientious on left and right, respectively. We can see that the changes made to the conversation scripts have improved the scores on all three sets.

C script, the language cues for 'vague' and 'negativity' were strengthened. For example, Low C's message, *"And what do you think about the stuff I told you about right now? Would it harm you to try it or was I being helpful?"* was given an opening line: *"Finally we continue.*

For High C personality, we increased the 'insightful' and 'informativeness' language cues. The result was a more proactive personality that gives the users more suggestions. For the High C literature self-help message, the part *"If you read at least a couple of books per month, that would be really beneficial for you, and if you read even more, then that would be great!"* was added.

We did not make any significant changes to the extroversion scripts. Some of the individual messages were often confused with High C, and the 'dissatisfaction' language cue was not paired with Low E as much as we had hoped, so we ended up making minor changes to that language cue. As an example, in *"Honestly, the best ways to deal with low mood are the ones you derive pleasure from. Unfortunately only you know what those are. What excites you? Try doing those things."* the ending was changed to *"... derive pleasure from. Don't ask me what that is, none knows those but you. Perhaps you could try to also talk to your friends if you enjoy that kind of thing?"*.

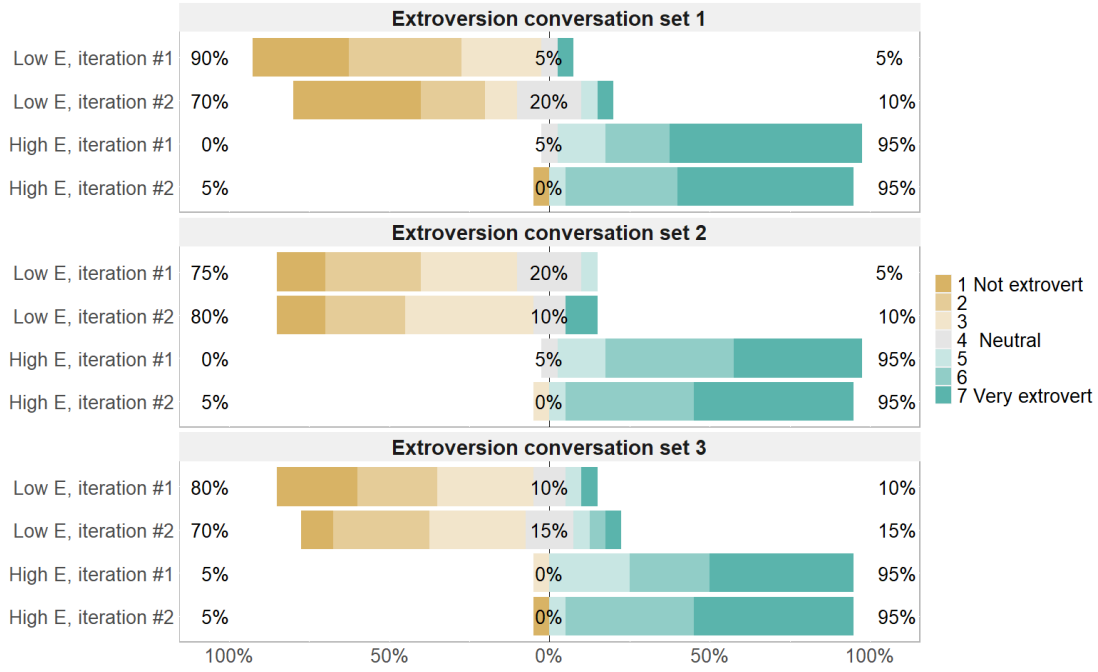

Figure 4: Results for the extroversion sets on the second section of the validation for both iterations. The figure shows the proportion of the participants rating the conversation set neutral (4) in the center, and low (1-3) and high (4-7) extroverted on left and right, respectively. No notable changes were reported for high extroversion, but low extroversion got worse results on the second iteration.

Table 2
Wilcoxon Rank Sum tests for differences between low and high variants of the evaluated personality traits (Conscientiousness, Extroversion).

	Set 1		Set 2		Set 3	
	W	p	W	p	W	p
Conscientiousness Iteration #1	14	<.005	11.5	<.005	12	<.0005
Conscientiousness Iteration #2	6.5	<.0005	3	<.0005	14	<.0005
Extroversion Iteration #1	3.5	<.0005	2	<.0005	4.5	<.0005
Extroversion Iteration #2	17	<.0005	11	<.0005	18	<.005

5.2. Second Iteration

After the changes, we conducted the second iteration of the validation. The changes made to the conversation script significantly improved the distinction between conscientiousness sets 1 and 2, as shown in Figure 3. Again, using the Wilcoxon Rank-Sum test there was a significant difference in set 2 ($p < .05$, $W = 125$) when comparing between iterations, but not for set 1 ($p = .30$, $W = 69$). Regardless, the visual inspection for set 1 indicated the separation between Low C and High C was more clear in iteration 2. Overall, the distinction between Low C and High C was now effectively communicated in the conversation script.

The language cues for extroversion showed some variance in accuracy compared to the first iteration but were still mostly correct. The only incorrect language cue was 'Realism' for extroversion set 1. The Likert-style items show the actual improvements, as even though Low E set 2 was rated higher on the extroversion scale (even up to scores of 7) than on iteration 1, the difference between Low E and High E was clear for all three question sets in both iterations (as can be seen in Figure 4).

The third section's results can be seen in Table 1. While the iteration did help with Low E, it was still often confused with High C.

Overall, after the second iteration, we can see significant improvements in the validation. We especially man-

aged to improve the distinction between the conscientiousness personality types.

5.3. Outlier Responses and Challenges of Big Five

The use of the Big Five characteristics has seen critique regarding their use for generalization, especially for more vague traits such as neuroticism. The conception of the traits can vary from person to person, even for the more obvious traits, such as those used in our design; extroversion and conscientiousness. 11 out of the 40 participants had one (or in some cases more than one) response which significantly differed from the general consensus, e.g., rating both options as high on a scale, or rating High C as Low C and vice versa. Clearly, in some cases, personal perception of what is considered, for example, conscientious behavior, can alter results in studies like ours. Big Five has been critiqued for its traits to be perceived differently according to different cultural upbringing or different language skills (first languages) for studies based on lexical analysis. However, we did not observe any significant influence of country of origin or first language on the outliers.

6. Discussion and Future Work

Our work aims to expand on previous research, e.g., Heudin et al. [12] showing increased performance of multi-personality CAs. Improving the user experience and making self-help tools more available could increase individuals' mental health well-being considerably. To pair the user with the CA best matching their desired traits and personality, we need to consider ways to match and adapt the CA depending on the user, which has been done for extroversion-based agents [27]. Our work focuses on the mental health self-help context and expands with consideration of the conscientiousness personality trait.

One of the key factors when designing personalities is consistency [26]. Identifying both low and high variants of conscientiousness and extroversion proved reliable. Cultural and language differences can lead to a different understanding of language cues, and especially for lexical analysis, the use of Big Five has received its share of critique [46]. Ultimately, one could ask, *"Do we know whose idea of conscientiousness has been encoded into this CA?"* - and from an individual's perspective, the language used by a CA could be accurate for one but misleading for others. A number of our Prolific workers stood out, but overall a vast majority were very much in agreement in assessing the CAs. However, careful consideration for cultural and language backgrounds should be considered in the future.

The next step for our generated CAs is to evaluate how their different personalities are perceived by individuals that are either seeking or could consider mental health help using self-help methods. We will also consider how the two scales correspond to the individual's own personality. Before moving on, we plan on validating the neutral conversation to make sure it is actually perceived as the neutral option on the two scales.

6.1. Limitations

The use of Big Five personality traits for analysis can be influenced by cultural or language-based differences between study participants. We did not observe any such differences according to the country of origin or first language of our study participants.

7. Conclusion

This paper presented the design and validation processes of the created CA personalities to be used in future studies. We found the low and high variants to be easily distinguishable from each other, but there are difficulties when choosing between extroversion and conscientiousness as several language cues are shared. Changes in the conversations improved the results. Further validation, including validation of the neutral personality, will be conducted before the next step in our research.

Acknowledgments

This research is connected to the GenZ strategic profiling project at the University of Oulu, supported by the Academy of Finland (project number 318930) and CRITICAL (Academy of Finland Strategic Research, 335729). Part of the work was also carried out with the support of Biocenter Oulu, spearhead project ICON.

References

[1] L. Laranjo, A. G. Dunn, H. L. Tong, A. B. Kocaballi, J. Chen, R. Bashir, D. Surian, B. Gallego, F. Magrabi, A. Y. Lau, et al., Conversational agents in healthcare: a systematic review, Journal of the American Medical Informatics Association 25 (2018) 1248–1258.

[2] J. L. Z. Montenegro, C. A. da Costa, R. da Rosa Righi, Survey of conversational agents in health, Expert Systems with Applications 129 (2019) 56–67.

[3] K. H. Ly, A.-M. Ly, G. Andersson, A fully automated conversational agent for promoting mental well-being: a pilot rct using mixed methods, Internet interventions 10 (2017) 39–46.

[4] A. N. Vaidyam, H. Wisniewski, J. D. Halamka, M. S. Kashavan, J. B. Torous, Chatbots and conversational agents in mental health: a review of the psychiatric landscape, The Canadian Journal of Psychiatry 64 (2019) 456–464.

[5] H. Gaffney, W. Mansell, S. Tai, et al., Conversational agents in the treatment of mental health problems: mixed-method systematic review, JMIR mental health 6 (2019) e14166.

[6] M. Mori, K. F. MacDorman, N. Kageki, The uncanny valley [from the field], IEEE Robotics & Automation Magazine 19 (2012) 98–100.

[7] A. Følstad, C. B. Nordheim, C. A. Bjørkli, What makes users trust a chatbot for customer service? an exploratory interview study, in: International conference on internet science, Springer, 2018, pp. 194–208.

[8] L. Ciechanowski, A. Przegalinska, M. Magnuski, P. Gloor, In the shades of the uncanny valley: An experimental study of human–chatbot interaction, Future Generation Computer Systems 92 (2019) 539–548.

[9] A. Muresan, H. Pohl, Chats with bots: balancing imitation and engagement, in: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems, 2019, pp. 1–6.

[10] R. Ahmad, D. Siemon, S. Robra-Bissantz, Communicating with machines: Conversational agents with personality and the role of extraversion, 2020. doi:10.24251/hicss.2021.492.

[11] A. Egges, S. Kshirsagar, N. Magnenat-Thalmann, A model for personality and emotion simulation, 2003. doi:10.1007/978-3-540-45224-9_63.

[12] J.-C. Heudin, Emotion selection in a multi-personality conversational agent, 2017. doi:10.5220/0006113600340041.

[13] H. Kim, D. Y. Koh, G. Lee, J.-M. Park, Y. kyung Lim, Designing personalities of conversational agents, CHI EA 39;19, ACM, May 02, 2019, pp. 1–6. URL: http://dl.acm.org/citation.cfm?id=3312887. doi:10.1145/3290607.3312887.

[14] F. Mairesse, M. A. Walker, M. R. Mehl, R. K. Moore, Using linguistic cues for the automatic recognition of personality in conversation and text, The Journal of artificial intelligence research 30 (2007) 457–500. URL: https://search.proquest.com/docview/2554117340. doi:10.1613/jair.2349.

[15] H. Qian, X. Li, H. Zhong, Y. Guo, Y. Ma, Y. Zhu, Z. Liu, Z. Dou, J.-R. Wen, Pchatbot: A large-scale dataset for personalized chatbot, ???? doi:10.1145/3404835.3463239.

[16] S. Castillo, P. Hahn, K. Legde, D. Cunningham, Personality analysis of embodied conversational agents, IVA 39;18, ACM, Nov 05, 2018, pp. 227–232. URL: http://dl.acm.org/citation.cfm?id=3267853. doi:10.1145/3267851.3267853.

[17] E. Ruane, S. Farrell, A. Ventresque, User Perception of Text-Based Chatbot Personality, Chatbot Research and Design, Springer International Publishing, Cham, 2021, pp. 32–47. URL: https://library.biblioboard.com/viewer/3af3b103-6756-11eb-8f43-0a9b31268bf5. doi:10.1007/978-3-030-68288-0_3.

[18] T. L. Smestad, F. Volden, Chatbot Personalities Matters, Internet Science, Springer International Publishing, Cham, 2019, pp. 170–181. URL: http://link.springer.com/10.1007/978-3-030-17705-8_15. doi:10.1007/978-3-030-17705-8_15.

[19] E. Svikhnushina, A. Placinta, P. Pu, User expectations of conversational chatbots based on online reviews, in: Designing Interactive Systems Conference 2021, 2021, pp. 1481–1491.

[20] C. Toxtli, A. Monroy-Hernández, J. Cranshaw, Understanding chatbot-mediated task management, in: Proceedings of the 2018 CHI conference on human factors in computing systems, 2018, pp. 1–6.

[21] B. De Raad, The big five personality factors: the psycholexical approach to personality., Hogrefe & Huber Publishers, 2000.

[22] G. Cameron, D. Cameron, G. Megaw, R. Bond, M. Mulvenna, S. O'Neill, C. Armour, M. McTear, Towards a chatbot for digital counselling, in: Proceedings of the 31st International BCS Human Computer Interaction Conference (HCI 2017) 31, 2017, pp. 1–7.

[23] Y.-C. Lee, N. Yamashita, Y. Huang, Designing a chatbot as a mediator for promoting deep self-disclosure to a real mental health professional, Proceedings of the ACM on Human-Computer Interaction 4 (2020) 1–27.

[24] K.-J. Oh, D. Lee, B. Ko, H.-J. Choi, A chatbot for psychiatric counseling in mental healthcare service based on emotional dialogue analysis and sentence generation, in: 2017 18th IEEE International Conference on Mobile Data Management (MDM), IEEE, 2017, pp. 371–375.

[25] D. Elmasri, A. Maeder, A Conversational Agent for an Online Mental Health Intervention, Brain Informatics and Health, Springer International Publishing, Cham, 2016, pp. 243–251. URL: http://link.springer.com/10.1007/978-3-319-47103-7_24. doi:10.1007/978-3-319-47103-7_24.

[26] C. Nass, K. M. Lee, Does computer-generated speech manifest personality? an experimental test of similarity-attraction, in: Proceedings of the SIGCHI conference on Human Factors in Computing Systems, 2000, pp. 329–336.

[27] A. Tapus, C. Ţăpuş, M. J. Matarić, User—robot personality matching and assistive robot behavior adaptation for post-stroke rehabilitation therapy,

Intelligent Service Robotics 1 (2008) 169–183.
[28] M. Komarraju, S. J. Karau, R. R. Schmeck, A. Avdic, The big five personality traits, learning styles, and academic achievement, Personality and individual differences 51 (2011) 472–477.
[29] J. Karsten, B. W. Penninx, H. Riese, J. Ormel, W. A. Nolen, C. A. Hartman, The state effect of depressive and anxiety disorders on big five personality traits, Journal of psychiatric research 46 (2012) 644–650.
[30] D. Peabody, L. R. Goldberg, Some determinants of factor structures from personality-trait descriptors., Journal of personality and social psychology 57 (1989) 552.
[31] L. R. Goldberg, An alternative" description of personality": the big-five factor structure., Journal of personality and social psychology 59 (1990) 1216.
[32] J. A. Kelly, J. M. Kern, B. G. Kirkley, J. N. Patterson, T. M. Keane, Reactions to assertive versus unassertive behavior: Differential effects for males and females and implications for assertiveness training, Behavior therapy 11 (1980) 670–682.
[33] D. M. Zeifman, Predicting adult responses to infant distress: Adult characteristics associated with perceptions, emotional reactions, and timing of intervention, Infant Mental Health Journal: Official Publication of The World Association for Infant Mental Health 24 (2003) 597–612.
[34] D. N. Klein, R. Kotov, S. J. Bufferd, Personality and depression: explanatory models and review of the evidence, Annual review of clinical psychology 7 (2011) 269–295.
[35] F. Mairesse, M. A. Walker, Towards personality-based user adaptation: psychologically informed stylistic language generation, User modeling and user-adapted interaction 20 (2010) 227–278. URL: https://link.springer.com/article/10.1007/s11257-010-9076-2. doi:10.1007/s11257-010-9076-2.
[36] S. Palan, C. Schitter, Prolific. ac—a subject pool for online experiments, Journal of Behavioral and Experimental Finance 17 (2018) 22–27.
[37] J. Pennebaker, L. King, Linguistic styles: Language use as an individual difference, Journal of personality and social psychology 77 (2000) 1296–312. doi:10.1037//0022-3514.77.6.1296.
[38] J.-M. Dewaele, A. Furnham, Extraversion: The unloved variable in applied linguistic research, Language Learning 49 (1999) 509–544. doi:https://doi.org/10.1111/0023-8333.00098.
[39] M. Mehl, S. Gosling, J. Pennebaker, Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life, Journal of personality and social psychology 90 (2006) 862–77. doi:10.1037/0022-3514.90.5.862.
[40] J. Weaver, Personality and self-perceptions about communication, Communication and personality: Trait perspectives (1998) 95–117.
[41] F. Heylighen, J.-M. Dewaele, Variation in the contextuality of language: An empirical measure, Foundations of science 7 (2002) 293–340.
[42] A. Furnham, Language and personality. (1990).
[43] S. Nowson, The language of weblogs: A study of genre and individual differences (2006).
[44] C. S. Cope, Linguistic structure and personality development., Journal of Counseling Psychology 16 (1969) 1.
[45] A. Thorne, The press of personality: A study of conversations between introverts and extraverts., Journal of Personality and Social Psychology 53 (1987) 718.
[46] J. Block, The five-factor framing of personality and beyond: Some ruminations, Psychological Inquiry 21 (2010) 2–25.

A. Detailed information on the language cues used

Table 3
Language cues used in the personality design, derived from the work of Mairesse [14]. Some of the language cues are shared between different personality traits.

Low Conscientiousness	High Conscientiousness	Low Extroversion	High Conscientiousness
Less perspective	More perspective	Single topic	Many topics, higher verbal output
Less careful	Checks that information is conveyed correctly	Realism	Exaggeration
More vague	Straight to the point	Problem talk	Pleasure talk
Few positive affect	Some positive affect	Dissatisfaction	Agreement and compliment
Many exclusive words (eg. but, without)	Few exclusive words	Non symphatetic	Sympathetic, concerned about heared
Many causation words	Few causation words	Eloborated constructions	Simple constructions
Few insight words	Many insight words	Few conjucations	Many conjucations
Many frequent words	Few frequent words	Many unfilled pauses	Few unfilled pauses
Impulsive	Not impulsive	Rich vocabulary	Poor vocabulary
Informal	Formal	Strict selection	Think out loud
Many references to friends	Few disfluencies, filler words	Formal language	Informal language
Many disfluencies, filler words	Few negations	Negative emotion words	Positive emotion words
Many negations	Few references to friends	Few words related to humans	Many words related to humans
Many swear words	Few swear words	Many uses of although	Few uses of although
Shorter words	Longer words	Many nouns, adjectives, prepositions	Many verbs, adverbs, pronouns
Many negative emotion words	Few negative emotion words	Many tentative words	Few tentative words
Few positive emotion words	Many positive emotion words	Many negations	Few negations
		Few swear words	Many swear words
		Longer words	Shorter words