






USE OF AI IN MHEALTH MOBILE APPLICATIONS: THE INFLUENCE OF VIRTUAL THERAPISTS' VISUAL CHARACTERISTICS ON USER PERCEPTION

Gala Golubović , Stefan Đurđević , Jelena Kerac , Iva Juretić , Sanja Cvetojević 

University of Novi Sad, Faculty of Technical Sciences,
Department of Graphic Engineering and Design, Novi Sad, Serbia

Abstract: Implementing artificial intelligence within the applications for mental health has improved their services. As the most common application, Chatbots that simulate communication with a therapist stand out. With this feature, individuals who shy away from accessing therapy due to the stigma associated with mental health have the opportunity to access therapy without fear of negative feedback. This research aimed to investigate the impact of virtual therapists' visual characteristics on users' reactions. Six characters, created using MetaHuman Creator software and augmented with audio recordings, were evaluated by respondents on parameters such as likability, naturalness of appearance, excitement, positivity, and sense of trustworthiness. Video stimuli were employed to encourage these assessments. Results indicated that the third female character received the highest likability rating, while the first male character was considered the most natural and reliable. The first female character aroused the most incredible excitement and positive feelings. In contrast, the second male character received the lowest ratings across multiple parameters, and the second female character scored poorly overall. Statistical analyses, including T-tests and one-factor ANOVA, revealed no significant variations across characters for most parameters except for excitement. Post hoc tests highlighted notable differences, especially between the second female character, other female characters, and the first male character. Correlation analysis revealed a strong positive relationship between likability and the naturalness of appearance with user trust. These findings underscore the importance of planning virtual therapists' visual characteristics to suit users, revealing interesting observations such as respondents' perceived differences in character voices and the greater attention to detail exhibited by male subjects towards male characters.

Key words: mHealth, mobile apps, virtual therapists, visual characteristics, user perception

1. INTRODUCTION

The use of medical mobile applications is extensive and varies depending on the field of medicine and its specific requirements. Consequently, the tasks and functions within the application also vary. This means that the functions of artificial intelligence implemented in these applications also differ. Some of the applications that stand out are used for contacting emergency medical help vehicles (Ghani et al., 2021), medical consultation (Sindhumol & Ardhet, 2019), public health promotion (Biswas, 2023), chronic disease monitoring (Neuhauser et al., 2013), monitoring of dermatological changes (Freeman et al., 2020), particular diseases diagnostics (Verde et al., 2019), physical rehabilitation (Lo et al., 2018), oral hygiene (Al-Jallad et al., 2022) and mental health.

Mental health is one of the fields of medicine in which communication and fast and easy contact with medical personnel are crucial. Since the patient's condition cannot be predicted, it is impossible to plan therapeutic sessions that will be carried out at the exact moment of relapse. Therefore, the availability of mobile applications specialised in mental health is a facilitating circumstance that can only benefit the therapeutic process. The application of artificial intelligence (AI) in these applications has extended their importance to a considerably higher level. Along with machine learning, artificial intelligence has significant potential to transform healthcare and improve efficiency and innovation in clinical practice (Agarwal et al., 2022). Milne-Ives et al. (2022) conducted a general review of the use of artificial intelligence in mobile applications for mental health. The most common uses are predicting various factors (stress, mood, risk), supporting conversations, providing diagnostic support in decision-making, personalised notifications, and recommending interventions. It was further highlighted that the applications addressed various mental health conditions such as depression, mood disorders, stress management, suicide risk, as well as diagnostic support. It was shown that this type of mobile application has no negative impact on its users. Results show that progress and improvement of the condition were noted, or there was no significant impact. The authors pointed out that the collection of indirect data and the direct information obtained from the user form a quality basis for planning the following steps (Milne-Ives et al., 2022). Gamble (2020) also reviewed the use of artificial intelligence in mental health

applications. The author points out that the most common implementation is through chatbots and as an additional resource during therapy with therapists. The chatbot can psycho-educate by recognising the user's feelings and finding material suitable for them. Additionally, it can offer mental health information anytime without analysing feelings. It was found that by monitoring user-approved activities or analysing their emotional status based on interactions, AI has the capacity to provide recommendations based on user preferences. As a result, content customisation can effectively motivate users, thereby encouraging constant engagement with the application. It was pointed out that by regularly monitoring mood data, artificial intelligence algorithms can create databases containing habits, moods, preferred coping mechanisms and other factors without fear of compromising the privacy and security of personal data. In addition, it was observed that the application of cognitive behavioural therapy (CBT) through applications is at a desirable level, with results that rival the application of live CBT techniques. In addition, diagnosing post-traumatic stress has proven to be very effective when using chatbots specialised for these purposes. Interestingly, certain age groups, such as adolescents and the younger population, avoid pursuing help when it is needed due to the stigma surrounding mental health. This is exactly where these mobile applications stand out. It is considerably more comfortable for this group of users to relax and venture into therapy through digital devices than in-person therapies. This shows that the benefits of applying artificial intelligence to mental health applications are fundamental. However, the author lists certain aspects that should be considered and improved. For example, there is a need for accountability, transparency, supervision, and the application of standards regarding chatbots so that ethical aspects are not questioned. Also, the target groups of users should be examined, and depending on their needs and preferences, the application should be designed in terms of both interface and functionality (Gamble, 2020). D'alfonso et al. (2017) wrote about using artificial intelligence through chatbots on a platform dedicated to mental health therapy in youth. The authors state that previous research indicates that users are more open and share information more efficiently, especially about sensitive topics, when they know they are interacting with a machine interface. It was pointed out that the benefit of using chatbots is reflected in the fact that an immediate psychological assessment and additional learning can be carried out based on the user's response in real time. It is indicated that this type of technology could be used to collect information about a patient before they come in contact with a professional. Based on the collected information, a preliminary profile of the patient could be formed. Additionally, with platforms intended for user-therapist communication, it has been suggested that the chatbot can be used for casual conversation while the patient waits for an appointment (D'alfonso et al., 2017). Berrouiguet et al. (2019), as well as Haines-Delmont et al. (2020), discussed the importance of applying artificial intelligence in mental health applications with implications for suicide prevention. Preliminary research has found that the application can identify changes in mobility patterns, which can be used as proxies for behavioural changes and mental health relapses (Berrouiguet et al., 2019). It has also been established that certain machine-learning algorithms can recognise real signs of suicidal desire and separate users who have this desire from those who do not (Haines-Delmont et al., 2020). Inkster, Sarda and Subramanian (2018) conducted a study on the effectiveness of the Wysa application in monitoring depressive conditions. Wysa is an AI-powered mobile application featuring emotional intelligence that aims to improve users' mental health using a text-based interface. It allows users to create a pleasant environment and adequately express their feelings. The results of the research showed that the conversations offered by the application have an impact on improving the symptoms of depression. A more significant impact was observed in users with more severe symptoms of depression. Users expressed a rather positive user experience with recommendations for improving the application. They expressed the desire for the application to understand them better and for the questions not to be asked repeatedly (Inkster, Sarda & Subramanian, 2018). Research conducted by Fitzpatrick, Darcy, and Vierhile (2017) aimed to evaluate the feasibility of delivering cognitive behavioural therapy via an artificial intelligence-based chatbot (Woebot). This software is designed to present appropriate cognitive-behavioural concepts to the user based on mood data through video content or word games that serve to educate the user about their current state. Functions that contributed to a positive user experience when using Woebot were the empathy shown by the chatbot, and its "personality"; encouraging learning was also highlighted as a positive side. Users mostly pointed out that a negative side of the software was the limitation of communication. They stated that they encountered a lack of understanding of certain topics by the chatbot or that confusion occurred if unexpected answers appeared. In addition, it has been observed that conversations can be repeated as completely identical (Fitzpatrick, Darcy & Vierhile, 2017). Ly, Ly, and Andersson (2017) tested the Shim chatbot. Shim is a fully automated agent (chatbot) created as an application whose user interface is similar to the interface of text messaging applications. The communications within the

application are based on strategies and activities aimed at positive psychology. The research showed that when using Shim, users often had the impression of communicating with a human being and developed a sense of some relationship. Users pointed out that the application also impacted their everyday lives, i.e., they began to apply the behaviour patterns suggested by the application. Additionally, as a positive side of the application, users pointed out that it was beneficial to have it with them at all times and to be able to access it. Also, the weekly status and progress reports were essential to the users so that they could follow their personal progress. On the negative side, they pointed out the repetition of content, the impossibility of entering deeper conversations, and that misunderstandings occur in certain situations. Based on all user responses, it is obvious that, at times, the application gives the feeling of a human being; however, when it continues to be used, there are disadvantages that artificial intelligence has compared to humans (Ly, Ly & Andersson, 2017). In 2020, Dosovitsky et al. researched the use of the Tess chatbot. This application offers numerous chatbot interaction modules that have proven to be a promising practice. Availability of modules led to a heterogeneous application usage pattern. The authors point out that focusing on developing short, simple, and consistent modules would be desirable, and testing them with small iterative studies would increase usability and user engagement (Dosovitsky et al., 2020). Fulmer et al. (2018) also examined the usability of this application. They determined that as many as 86% of respondents were satisfied when using the application, while 80% gained new knowledge during use. As the best aspects of the application, users highlighted responsibility and accessibility, the empathy shown by the chatbot, and the accession to new knowledge initiated by the bot. As negative aspects, users cited limitations like conversations and limited understanding. Most importantly, a significant reduction in anxiety symptoms was observed in users who continuously used the application. They showed a higher level of engagement and overall satisfaction and emphasized that the therapeutic experience was pleasant (Fulmer et al., 2018). Jiménez-Serrano, Tortajada, and García-Gómez (2015) investigated the ability of a machine learning-based application model to predict symptoms of postpartum depression. The application is based on a simple questionnaire on which it is determined whether a new mother is at risk of developing postpartum depression. The obtained results enable early diagnosis and intervention if the need arises. It was found that this type of application can provide early detection of postpartum depression, that the results are easy to interpret, and that the approach is cost-effective. This application has not been clinically evaluated, and a proposal was made to implement it to be applied in actual practice (Jiménez-Serrano, Tortajada & García-Gómez, 2015).

The newest implementation of AI in mental health applications is through virtual humans. Ellie represents one of the most interesting implementations of artificial intelligence within mental health applications. Ellie is a virtual human who conducts semi-structured conversations aimed at creating an interaction that will enable the automatic assessment of stress indicators, defined through verbal and non-verbal behaviours related to depression, anxiety or post-traumatic stress disorder. It is designed to follow the micro-expressions of the person with whom it communicates, react to facial expressions and gestures, perform compassionate gestures, and, therefore, build a more pleasant relationship with the patient. Some research has shown that Ellie can recognize subconscious physical cues much better than many highly trained psychologists. She is designed to recognize 60 non-verbal cues per second, everything from eye gaze to facial tilt to tone of voice. These abilities have proven to be highly significant and effective in diagnosing mental illnesses. It was also observed that users who interacted with Ellie would rather interact with her again than with an actual therapist. Experts believe this is due to the absence of judgment in communication with Ellie (Pearce, 2020). The research conducted by Rizzo et al. (2016) showed that the subjects felt comfortable communicating with Ellie and shared intimate information without problems during the conversation. However, respondents pointed out that Ellie did not adequately recognize their nonverbal behaviour and that her body language and nonverbal behaviours were often inappropriate (Rizzo et al., 2016). Ellie is yet to be available for public use. Through their research, Philip et al. (2017) examined the abilities of a virtual human (Julia) to diagnose depressive disorders with a comparison to the diagnosis of a clinical psychiatrist. In this research, it was determined that Julia showed satisfactory results when it came to diagnosing depressive conditions. Although the diagnostic efficiency was at a high level, it was observed that there were deficiencies in patients with milder symptoms. It was additionally pointed out that it would be essential to enable her to recognize non-verbal aspects, such as facial expressions and changes in the tone of her voice, which could lead to exceptional progress in psychiatry. The authors believe using artificial intelligence through virtual humans could significantly save doctors time during diagnosis and provide patients with a completely unbiased approach based on a theoretical diagnosis model (Philip et al., 2017).

2. BACKGROUND

The literature review reveals that the predominant utilisation of artificial intelligence within mental health applications is through the integration of chatbots. This type of chatbot primarily aims to replicate interactions with a therapist, typically through written exchanges. However, a relevant concern arises regarding the deficiency in emotional intelligence exhibited by these chatbots. In addition to this application of artificial intelligence, the question arises about what functions it could fulfil while positively influencing users and allowing them to free themselves from inner turmoil. Some potential possibilities for implementing artificial intelligence within applications intended for mental health are mood monitoring, personalised recommendations, virtual therapist, recognition of emotions, analysis of feelings, user engagement and gamification, monitoring medication therapy, and early diagnosis and risk assessment. Considering the insights into both present and prospective utilisation of artificial intelligence in mental health applications, the concept of a virtual therapist is particularly interesting.

3. METHODS

The research is based on analysing user responses to virtual characters whose role is to simulate a virtual psychotherapist. A survey served as the fundamental tool for documenting participants' reactions.

3.1 Characters

Models of virtual humans were used as stimuli, presented as six characters - three female and three male characters (Figure 1).



Figure 1: Characters

These characters were designed to exhibit maximal diversity, achieved through facial features, skin tones, hair, and eye colours. MetaHuman Creator software generated 3D character models (Figure 2). Initially, a selection of pre-existing models served as the foundation for character development, with subsequent adjustments to align with the predetermined appearance criteria. Regarding the character's surroundings, a studio setting with a neutral grey backdrop was chosen to minimise potential distractions for participants.

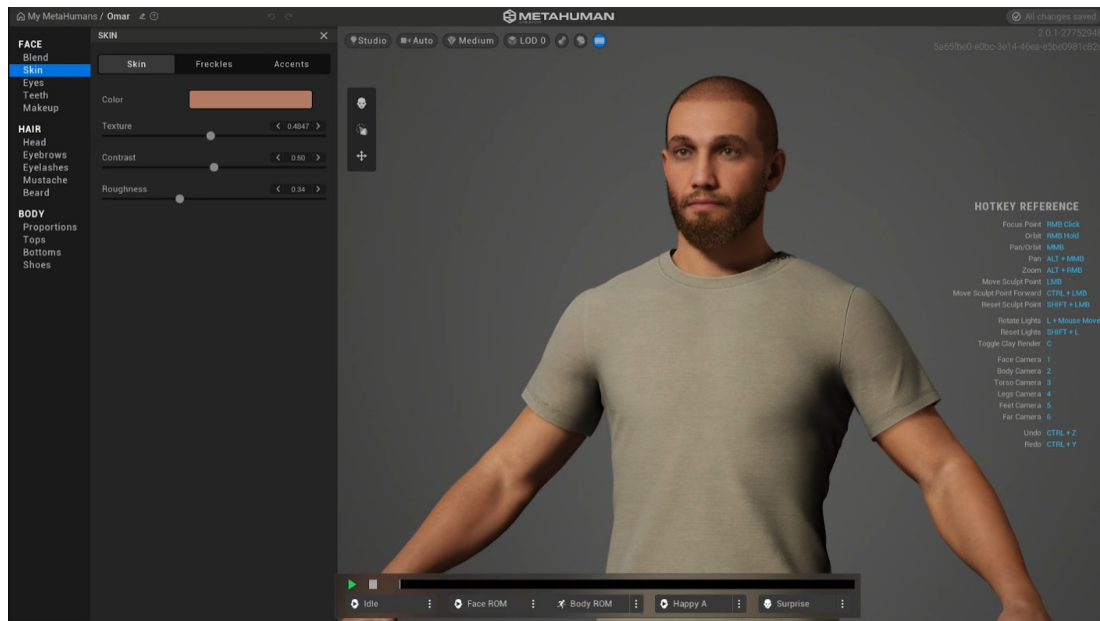


Figure 2: The work environment of the MetaHuman Creator software

In addition to character development, voice content was recorded to simulate introductory communication with a virtual therapist. Two audio recordings, one featuring a male voice and the other a female voice, were utilised to ensure compatibility with characters of both genders. Employing a unique voice for each gender aimed to direct attention solely towards the visual attributes of the character, thus ensuring that only the physical appearance influenced respondents' reactions. The audio recordings were based on a predefined script, which simulated a welcoming and familiarising interaction with the application. The D-ID AI Video Generator was used to obtain the final video, seamlessly integrating visual and audio components, which allowed the characters to synchronise their gestures with the spoken content from the audio. The final stimuli were videos that included generated characters accompanied by audio content. The videos are generated to last 45 seconds each.

3.2 Examination process

In addition to the previously mentioned videos, a survey was also used for the examination. The survey encompassed seven sections, with the initial section addressing general respondent information, including gender, prior utilisation of mental health applications, and familiarity with technologies such as virtual humans. Subsequently, the following six sections comprised six questions, individually assessing aspects of each character. These included the likability of the characters, the perceived naturalness and trustworthiness of their appearance, the level of enthusiasm and positivity experienced during communication, and the inclination to sustain communication with the portrayed therapist.

4. RESULTS AND DISCUSSION

Given that the survey comprises two distinct types of sections, the subsequent analysis of results will adhere to the same principle.

4.1 Analysis of the general part of the survey

Twenty participants were engaged in the study, comprising an equal split of ten females and ten males aged between 25 and 45. The respondent group is characterised by homogeneity in terms of gender, age, and educational background, with each individual having attained some level of higher education, ranging from undergraduate to doctoral studies. Moreover, all participants exhibit familiarity with the concept of virtual humans or analogous technologies (80% of them). Notably, only a minority of the respondents, representing 10% of the group (2 individuals), have previously utilised mental health applications.

4.2 Analysis of the respondents' reaction to the presented characters

The characters were evaluated based on the following parameters:

- likeability,
- naturalness of appearance,
- trustworthiness,
- respondents' excitement,
- inducing a positive/negative feeling in the respondents and
- desires of respondents to continue the conversation with them.

4.2.1 First male character

The first male character is generated to possess dark attributes, including hair and beard colour, eye hue, and complexion. Respondents frequently assigned a likability rating of 4 out of 5 to this character, with an average rating of 3.8. Similarly, its appearance was predominantly rated as natural, with a typical score of 4 and an average rating of 3.95. Notably, observations indicated that these ratings remained consistent across genders. Additionally, it was observed that the ratings for likability and naturalness of appearance did not necessarily correlate proportionally. The parameter of causing excitement among users was predominantly rated 6 out of 9, resulting in an average score of 4.55. Conversely, the sensation of positivity garnered higher ratings, averaging 6.6 out of 9. Regarding trustworthiness, respondents provided an average rating of 4 on a scale from 1 to 5 for this character. Interestingly, a slightly elevated average rating was observed among female respondents compared to their male counterparts. Merely 3 participants, constituting 15% of the group, indicated that they would not pursue further communication with the depicted character. There is a noticeable correlation between this determination regarding continued communication and the overall ratings across all parameters. The generally positive attitude conveyed by respondents towards this character interested them and awakened the desire to explore the possibilities of this character further.

4.2.2 Second male character

The second male character was generated to distinguish itself from the first through fundamental traits. Differences are primarily evident in facial features, hair colour and texture, complexion, hairstyle, and beard/moustache configuration. The likability of this character received an average rating of 2.65 on a scale ranging from 1 to 5, while the perceived naturalness of its appearance garnered an average rating of 3.1. Respondents expressed unfavourable impressions of this character, noting that its physical attributes evoked negative associations and even elicited fear. The excitement generated by this character among respondents was assessed with an average rating of only 4.1 on a scale from 1 to 9. Similarly, the character's capacity to evoke positivity in users yielded an average rating of 4.8 out of 9. Trustworthiness, as perceived by users, garnered an average score of 2.8 on a scale of 1 to 5. Notably, trust levels were observed to correlate directly with likability and the perceived naturalness of appearance, with respondents assigning similar ratings to these three parameters. Despite the character receiving low average ratings for trustworthiness, 45% of respondents expressed their intention to continue communicating with it.

4.2.3 Third male character

The final male character was generated to diverge from both preceding ones. Similar to earlier, the primary distinctions refer to hair colour, eye hue, complexion, and hair and beard style. Respondents noted, among other observations, that this character possesses softer facial features, evoking associations with a medical professional. Conversely, some respondents expressed that the specific characteristics of this character repelled them, as they triggered associations related to criminality. The likability of this character garnered an average rating of 3.75 on a scale of 1 to 5. Similarly, the perceived naturalness of its physical appearance received a comparable rating, averaging 3.85 out of 5. Notably, female respondents gave slightly higher ratings than their male counterparts. The level of excitement generated by this character among respondents was generally neutral, with an average score of 4.8 on a scale from 1 to 9. However, a slightly more positive sentiment was observed regarding the character's ability to evoke positive feelings, with an average rating of 6.6 out of 9. It was noted that the ratings provided by female respondents were marginally higher than those of male respondents in this aspect.

Overall, respondents expressed a predominantly positive perception of trustworthiness associated with this character, with an average rating of 3.55 out of 5. Once again, evaluations from female respondents tended to be slightly more optimistic. Impressively, 70% of respondents indicated their willingness to continue communication with this character, with a more significant proportion being female respondents.

4.2.4 First female character

Like the initial male character, the first female character is designed to exhibit dark attributes such as hair colour, eye hue, and complexion. The likability of this character received a notably high average rating of 4.15 out of 5. It was noted that male respondents tended to rate the character somewhat less favourably than female respondents. Conversely, the perceived naturalness of the character's appearance garnered a slightly lower average score of 3.7 out of 5, with male respondents assigning higher ratings than female respondents. The excitement induced by the character was rated at 4.95 on a scale from 1 to 9. Additionally, positivity was evaluated with an average score of 6.9 out of 9, with female respondents displaying a more positive inclination towards the character. Users' trust in this character was assessed with an average score of 3.95 out of 5, with male respondents assigning slightly lower ratings than female respondents. Regarding communication with this character, most respondents, accounting for 90% of cases, expressed their intention to continue using the application if it involved interaction with this character.

4.2.5 Second female character

When generating another female character, the attributes were selected to align with an individual's standard with fair skin, hair, and eyes. Regarding likability, this character received the lowest average rating among all female characters, with an average score of 2.85 on a scale ranging from 1 to 5. Similarly, the naturalness of her appearance also garnered the lowest average rating among all evaluated characters, with an average score of 2.7 out of 5. Additionally, the level of excitement experienced by users during interaction with this character was rated the lowest across all characters, with an average score of 3.9 on a scale from 1 to 9. It is apparent that, on average, male respondents reacted more negatively to this character in terms of excitement. The impact on the feeling of positivity was rated as neutral, with an average score of 5 on a scale from 1 to 9. However, female subjects exhibited significantly more positive reactions to this parameter than male subjects. The parameter of trustworthiness was also rated relatively low, with an average score of 2.5 out of 5, marking the lowest rating among all characters. Responses for this parameter were slightly more positive among female respondents than male respondents. Despite evaluating all parameters relatively low, 40% of respondents expressed their intention to continue communicating with this character. Of these, the distribution between male and female respondents is equal, with 20% each (4 respondents).

4.2.6 Third female character

The final character, a female, was crafted with gentle facial features, light eyes, and hair of non-natural colour to maximise differentiation from the preceding characters. The likability of the final character accumulated a notably high average rating of 4.2 out of 5, making it the highest-rated character in terms of likability among all characters. Respondents consistently assigned ratings from 3 to 5 when assessing this parameter, indicating an overwhelmingly positive overall impression. Additionally, the naturalness of the character's physical appearance received a commendable rating, averaging 3.6 out of 5. Interestingly, this parameter received higher average ratings from male respondents. The overall level of excitement induced by the character was perceived as neutral, with an average rating of 5.2 on a scale from 1 to 9. Similarly, the ratings provided by male respondents were slightly higher than those of female respondents for this parameter. With such an average, this character emerges as the best-rated in terms of respondents' arousal levels. The extent of positivity evoked by this character was evaluated with an average score of 6.75 on a scale from 1 to 9. In this aspect, the ratings given by female respondents were slightly higher. The level of trust obtained by the character was predominantly rated as 4, resulting in an average rating of 3.85 out of 5. Notably, female respondents expressed a more positive reaction compared to male respondents in this regard. Remarkably, 80% of respondents expressed their intention to continue communication with this character, aligning with the expected outcome based on all other parameters considered.

4.3 Statistical analysis of the results

A statistical analysis of the acquired results was conducted to ascertain the impact of individual parameters on user perception and overall impression. This analysis employed statistical methodologies such as the one-factor analysis of variance (ANOVA) and T-test to compare groups. At the same time, correlation served as a statistical technique for establishing relationships between variables.

4.3.1 Individual character analysis

To examine the impact of respondent gender on the outcomes, an independent analysis of the results for each evaluated parameter of every character was conducted. Using the T-test, comparisons were made regarding the parameters of likability, naturalness of character appearance, excitement and positivity elicited from respondents, and the level of trustworthiness perceived by respondents during interaction with the character.

The T-test results of independent samples were employed to compare the test outcomes of parameters including likability, naturalness of character appearance, excitement, positivity, and respondents' perception of trustworthiness. The analysis did not disclose a statistically significant variation in the responses between female and male respondents across all investigated parameters for all six characters.

4.3.2 Analysis of individual parameters

Following the determination that gender does not statistically influence the outcomes for all six characters, further analyses were conducted encompassing all respondents. A statistical analysis was performed utilising the one-factor ANOVA to explore variations in parameter values contingent upon the character. The test was executed by summing up all five parameters' values. This procedure tested for deviations within each parameter across all characters (Table 1).

Table 1: Results of the One-Way ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Likeability	Between Groups	44.175	5	8.835	9.579	0.000
	Within Groups	105.150	114	0.922		
	Total	149.325	119			
Naturalness of appearance	Between Groups	23.467	5	4.693	4.842	0.000
	Within Groups	110.500	114	0.969		
	Total	133.967	119			
Excitement	Between Groups	25.267	5	5.053	1.121	0.353
	Within Groups	513.900	114	4.508		
	Total	539.167	119			
Positivity	Between Groups	89.242	5	17.848	5.615	0.000
	Within Groups	362.350	114	3.179		
	Total	451.592	119			
Trust	Between Groups	40.942	5	8.188	6.196	0.000
	Within Groups	150.650	114	1.321		
	Total	191.592	119			

The conducted ANOVA yielded results indicating the presence of variations in the values attributed by respondents to the parameters, which exhibit exceptional statistical significance. The sole parameter where variations were not statistically significant is arousal. For all other parameters, the maximum statistical significance value (<0.0005) was observed.

Since significant statistical differences were identified regarding the likability and naturalness of the characters and the positivity and trustworthiness perceived by the respondents, post-hoc tests were administered to ascertain which characters exhibited notable statistical differences (Table 2).

Table 2: The influence of individual characters on test parameters

Dependent Variables	(I) Character	(J) Character	Mean Difference (I-J)	Std. Error	Sig.
Likeability	Male character 1	Male character 2	1.200*	0.304	0.002
		Female character 2	1.000*	0.304	0.016
	Male character 2	Male character 3	-1.100*	0.304	0.006
		Female character 1	-1.500*	0.304	0.000
		Female character 3	-1.550*	0.304	0.000
	Male character 3	Female character 2	0.900*	0.304	0.042
	Female character 1	Female character 2	1.300*	0.304	0.001
	Female character 2	Female character 3	-1.350*	0.304	0.000
Naturalness of appearance	Female character 2	Male character 1	-1.250*	0.311	0.001
		Male character 3	-1.150*	0.311	0.004
		Female character 1	-1.000*	0.311	0.021
Positivity	Male character 2	Male character 1	-1.800*	0.564	0.022
		Male character 3	-1.800*	0.564	0.022
		Female character 1	-2.100*	0.564	0.004
		Female character 3	-1.950*	0.564	0.010
	Female character 2	Female character 1	-1.900*	0.564	0.013
		Female character 3	-1.750*	0.564	0.028
Trustworthiness	Male character 1	Male character 2	1.200*	0.364	0.016
		Female character 2	1.500*	0.364	0.001
	Male character 2	Female character 1	-1.150*	0.364	0.024
		Female character 3	-1.050	0.364	0.051
	Female character 1	Female character 2	1.450*	0.364	0.002
	Female character 2	Female character 3	-1.350*	0.364	0.004

Through a one-factor analysis of variance, the impact of the character models on parameters such as likability, naturalness of appearance, excitement, positive attitude of the respondents, and trust was investigated. Given that six characters were generated, the analysis involved comparing these six groups. A significant statistical difference was observed across all tested parameters, most prominently when contrasting male character 2 with male characters 1 and 3 and female characters 1 and 3. As for female character 2, a notable statistical difference was only detected when compared with male character 2. The most pronounced statistical significance of difference was noted in the likability parameter, particularly between male character 2 and female characters 1 and 3, and when comparing female character 2 with female character 3. The mean values of the differences are presented with either a positive or negative sign based on the direction of the comparison. For instance, when evaluating the likability expressed by male character one and male character 2, the mean difference is 1.2. These findings suggest that the ratings for the likability parameter of male character 1 are higher than those given for male character 2. Conversely, the scenario is reversed when examining the comparison of likability for male character two and female character 1. The mean difference value is 1.5, indicating that, in this case, the ratings of the first character (male 2) are lower than those of the character with which it is being compared (female 1). This interpretation of the results aligns with the ratings observed in the descriptive analysis.

4.3.3 Determination of correlation between parameters

A statistical correlation test was conducted to assess the impact of the likability and naturalness of the character's appearance on user trust. This analytical approach provides insight into the strength of the relationship between the parameters under examination.

Upon analysing the correlation between the parameters of likability and trustworthiness, as well as the naturalness of the appearance and respondents' trust, it was determined that a positive correlation exists between these parameters, exhibiting exceptional statistical significance (Table 3). This positive correlation signifies that the parameters are directly proportional, providing a clear understanding that the more likeable and natural a character is, the higher the level of trust. Regarding the strength of the relationship between the parameters, it was determined that both pairs of parameters were of high strength.

Table 3: Correlation test results

		Trustworthiness
Likeability	Pearson Correlation	0.766**
	Sig. (2-tailed)	0.000
	N	120
Naturalness of appearance	Pearson Correlation	0.600**
	Sig. (2-tailed)	0.000
	N	120

Calculating the coefficient of determination offers insight into the percentage of variance of one parameter caused by the variance of another parameter. In the case of the first pair of parameters tested (likability and trustworthiness), this coefficient stands at 59%. In contrast, in the case of the second pair (naturalness of appearance and trustworthiness) is somewhat lower at 36%. Consequently, likability accounts for 59% of the variance in responses concerning trustworthiness, while naturalness explains 36% of the variance in trustworthiness.

5. CONCLUSIONS

This research aimed to investigate the impact of virtual therapists' visual factors on users' reactions. Respondents' responses to six characters, created using MetaHuman Creator software and supplemented with audio recordings, were examined. Parameters such as the character's likability and naturalness of appearance, user excitement and positivity, and the sense of trustworthiness evoked by the characters were evaluated. Video stimuli were employed.

Based on respondents' subjective assessments, female character 3 received the highest likability rating, while male character 1 was considered the most natural and reliable. Female character 1 produced the greatest excitement and positive feelings among respondents. Conversely, male character 2 received the lowest ratings regarding likability, excitement, and positivity, while female character 2 scored poorly across all parameters. Overall, female character 1 was perceived most positively, while female character 2 garnered the most negative perception. Statistical analyses were conducted to assess variations in respondents' responses for each character individually. The T-test results indicated no statistically significant variations for any parameter across all characters. Additionally, one-factor ANOVA revealed significance only in the excitement parameter. Post hoc tests identified notable differences, particularly between female character 2, the other two female characters, and male character 1. Furthermore, a correlation analysis was conducted to explore the relationship between likability and the naturalness of the character with user trust. Both pairs exhibited a strong positive correlation.

Based on the results, it is concluded that when creating characters representing virtual therapists, it is necessary to take care of their visual characteristics, and adapt them to the user. Interestingly, some respondents perceived character voices as different, even though only one per gender was used. In addition, it was observed that male respondents notice more details when observing male characters and potentially compare themselves to them. In order to carry out the implementation of this type of content within mobile applications for mental health, it would be necessary to expand the base of examined characters with more intense mutual differences and include a more significant number of respondents. It would be interesting to examine the influence of the colour of the voice, the clothing and the environment in which the character is located on the respondents' reactions.

6. ACKNOWLEDGMENTS

This research has been supported by the Ministry of Science, Technological Development and Innovation (Contract No. 451-03-65/2024-03/200156) and the Faculty of Technical Sciences, University of Novi Sad through project "Scientific and Artistic Research Work of Researchers in Teaching and Associate Positions at the Faculty of Technical Sciences, University of Novi Sad" (No. 01-3394/1).

7. REFERENCES

- Agarwal, S., Jalan, M., Wilcox, H.C., Sharma, R., Hill, R., Pantalone, E., Thrul, J., Rainey, J.C. & Robinson, K.A. (2022) Evaluation of mental health mobile applications. *Technical Brief*. 41. Rockville (MD): Agency for Healthcare Research and Quality (US).
- Al-Jallad, N., Ly-Mapes, O., Hao, P., Ruan, J., Ramesh, A., Luo, J., Wu, T.T., Dye, T., Rashwan, N., Ren, J. & Jang, H. (2022) Artificial intelligence-powered smartphone application, AICaries, improves at-home dental caries screening in children: Moderated and unmoderated usability test. *PLOS digital health*. 1 (6), e0000046. Available from: doi: 10.1371/journal.pdig.0000046
- Berrouiguet, S., Barrigón, M.L., Castroman, J.L., Courtet, P., Artés-Rodríguez, A. & Baca-García, E. (2019) Combining mobile-health (mHealth) and artificial intelligence (AI) methods to avoid suicide attempts: the Smartcrises study protocol. *BMC psychiatry*. 19 (1), 1-9. Available from: doi: 10.1186/s12888-019-2260-y
- Biswas, S.S. (2023) Role of Chat GPT in Public Health. *Annals of biomedical engineering*. 51 (5), 868-869. Available from: doi: 10.1007/s10439-023-03172-7
- D'alfonso, S., Santesteban-Echarri, O., Rice, S., Wadley, G., Lederman, R., Miles, C., Gleeson, J. & Alvarez-Jimenez, M. (2017) Artificial intelligence-assisted online social therapy for youth mental health. *Frontiers in psychology*. 8, 796. Available from: doi: 10.3389/fpsyg.2017.00796
- Dosovitsky, G., Pineda, B.S., Jacobson, N.C., Chang, C. & Bunge, E.L. (2020) Artificial intelligence chatbot for depression: descriptive study of usage. *JMIR Formative Research*. 4 (11), e17065. Available from: doi: 10.2196/17065
- Fitzpatrick, K.K., Darcy, A. & Vierhile, M. (2017) Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. *JMIR mental health*. 4 (2), e7785. Available from: doi: 10.2196/mental.7785
- Freeman, K., Dinnes, J., Chuchu, N., Takwoingi, Y., Bayliss, S.E., Matin, R.N., Jain, A., Walter, F.M., Williams, H.C. & Deeks, J.J. (2020) Algorithm based smartphone apps to assess risk of skin cancer in adults: systematic review of diagnostic accuracy studies. *BMJ*. 368. Available from: doi: 10.1136/bmj.m127
- Fulmer, R., Joerin, A., Gentile, B., Lakerink, L. & Rauws, M. (2018) Using psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: randomized controlled trial. *JMIR mental health*. 5 (4), e9782. Available from: doi: 10.2196/mental.9782
- Gamble, A. (2020) Artificial intelligence and mobile apps for mental healthcare: a social informatics perspective. *Aslib Journal of Information Management*. 72 (4), 509-523. Available from: doi: 10.1108/AJIM-11-2019-0316
- Ghani, T., Jahan, N., Rasheduzzaman, M., Marzan, Y. & Khan, M.M. (2021) Design and feasibility analysis of an artificial intelligence based mobile app for emergency ambulance. In *2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC), NV, USA*. pp. 1147-1150. Available from: doi: 10.1109/CCWC51732.2021.9375907
- Haines-Delmont, A., Chahal, G., Bruen, A.J., Wall, A., Khan, C.T., Sadashiv, R. & Fearnley, D. (2020) Testing suicide risk prediction algorithms using phone measurements with patients in acute mental health settings: feasibility study. *JMIR mHealth and uHealth*. 8 (6), e15901. Available from: doi: 10.2196/15901
- Inkster, B., Sarda, S. & Subramanian, V. (2018) An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: real-world data evaluation mixed-methods study. *JMIR mHealth and uHealth*. 6 (11), e12106. Available from: doi: 10.2196/12106
- Jiménez-Serrano, S., Tortajada, S. & García-Gómez, J.M. (2015) A mobile health application to predict postpartum depression based on machine learning. *Telemedicine and e-Health*. 21 (7), 567-574. Available from: doi: 10.1089/tmj.2014.0113
- Lo, W., Lei, D., Li, L., Huang, D. & Tong, K. (2018) The Perceived Benefits of an Artificial Intelligence - Embedded Mobile App Implementing Evidence-Based Guidelines for the Self-Management of Chronic Neck and Back Pain: Observational Study. *JMIR Mhealth and Uhealth*. 6 (11), e8127. Available from: doi: 10.2196/mhealth.8127

- Ly, K.H., Ly, A.M. & Andersson, G. (2017) A fully automated conversational agent for promoting mental well-being: A pilot RCT using mixed methods. *Internet interventions*. 10, 39-46. Available from: doi: 10.1016/j.invent.2017.10.002
- Milne-Ives, M., Selby, E., Inkster, B., Lam, C. & Meinert, E. (2022) Artificial intelligence and machine learning in mobile apps for mental health: A scoping review. *PLOS Digital Health*. 1 (8), e0000079. Available from: doi: 10.1371/journal.pdig.0000079
- Neuhauser, L., Kreps, G.L., Morrison, K., Athanasoulis, M., Kirienko, N. & Van Brunt, D. (2013) Using design science and artificial intelligence to improve health communication: ChronologyMD case example. *Patient education and counselling*. 92 (2), 211-217. Available from: doi: 10.1016/j.pec.2013.04.006
- Pearce, K. (2020) AI Therapy: Watch a demo of the world's first virtual therapist. Available from: <https://www.diygenius.com/ai-therapy/> [Accessed 8th august 2023]
- Philip, P., Micoulaud-Franchi, J. A., Sagaspe, P., Sevin, E. D., Olive, J., Bioulac, S. & Sauteraud, A. (2017) Virtual human as a new diagnostic tool, a proof of concept study in the field of major depressive disorders. *Scientific reports*. 7 (1), 42656. Available from: doi: 10.1038/srep42656
- Rizzo, A., Scherer, S., DeVault, D., Gratch, J., Artstein, R., Hartholt, A., Lucas, G., Marsella, S., Morbini, F., Nazarian, A. & Stratou, G. (2016) Detection and computational analysis of psychological signals using a virtual human interviewing agent. *Journal of Pain Management*. 9 (3), 311-321.
- Sindhumol, S. & Ardhet, A. (2019) Artificial Intelligence Based Android App for Medical Consultation. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*. 8 (9), 1840-1846. Available from: doi: 10.35940/ijitee.I8229.078919
- Verde, L., De Pietro, G., Alrashoud, M., Ghoneim, A., Al-Mutib, K.N. & Sannino, G. (2019) Leveraging artificial intelligence to improve voice disorder identification through the use of a reliable mobile app. *IEEE Access*. 7, 124048-124054. Available from: doi: 10.1109/ACCESS.2019.2938265

