A multilingual digital mental health and wellbeing chatbot (ChatPal): pre-post multicenter intervention study

Potts, C., Lindstrom, F., Bond, RR., Mulvenna, M., Booth, F., Ennis, E., Parding, K., Kostenius, C., Broderick, T., Boyd, K., Vartiainen, A-K., Nieminen, H., Burns, C., Bickerdike, A., Kousmanen, L., Dhanapala, I. S. A., Vakaloudis, A., Cahill, B., MacInnes, M., ... O'Neill, S. (Accepted/In press). A multilingual digital mental health and wellbeing chatbot (ChatPal): pre-post multicenter intervention study. *Journal of Medical Internet Research*.

Link to publication record in Ulster University Research Portal

Document Version
Author Accepted version

A multilingual digital mental health and wellbeing chatbot (ChatPal): pre-post multicenter intervention study

Abstract

Background: In recent years, advances in technology have led to an influx of mental health apps, in particular the development of mental health and wellbeing chatbots which have already shown promise in terms of their efficacy, availability and accessibility. The ChatPal chatbot was developed to promote positive mental wellbeing amongst citizens living in rural areas. ChatPal is a multilingual chatbot, available in English, Scottish Gaelic, Swedish and Finnish, containing psychoeducational content and exercises such as mindfulness and breathing, mood logging, gratitude, and thoughts diaries.

Objective: The primary objective of this study is evaluate a multilingual mental health and wellbeing chatbot (ChatPal) to establish if it has an effect on mental wellbeing. Secondary objectives include investigating the characteristics of individuals that showed improvements in wellbeing along with those with worsening wellbeing and applying thematic analysis to user feedback.

Methods: A pre-post intervention study was conducted where participants were recruited to use the intervention (ChatPal) for a 12-week period. Recruitment took place across five regions: Northern Ireland, Scotland, Republic of Ireland, Sweden and Finland. Outcome measures included the Short Warwick Edinburgh Mental Wellbeing Scale, World Health Organisation-Five Well-Being Index, and Satisfaction with Life Scale were evaluated at baseline, midpoint, and endpoint. Written feedback was collected from participants and subjected to qualitative analysis to identify themes.

Results: A total of 348 people were recruited to the study (75% female, 25% male) aged between 18-73 (mean age 30). The wellbeing scores of participants improved from baseline to midpoint and from baseline to end-point, however improvement in scores were not statistically significant on the Short Warwick Edinburgh Mental Wellbeing Scale (P=.42), World Health Organisation-Five Well-Being Index (P=.52), or Satisfaction with Life Scale (P=.81). Individuals that had improved wellbeing scores (n=16) interacted more with the chatbot and were significantly younger compared to those whose wellbeing declined over the study (P=.025). Three themes were identified from user feedback including 'positive experiences', 'mixed/ neutral experiences' and 'negative experiences'. Positive experiences included enjoying exercises provided by the chatbot, while most of the mixed/ neutral or negative experiences mentioned liking the chatbot overall, but that there were some barriers, such as technical or performance errors which needed to be overcome.

Conclusions: Marginal improvements in mental wellbeing were seen in those who used ChatPal, albeit non-significant. We propose that the chatbot could be used along with other service offerings to complement different digital or face-to-face services, although further research should be carried out to confirm the effectiveness of this approach. Nonetheless, this paper highlights the need for blended service offerings in mental health care.

Introduction

On a global scale, the need for mental health related services has been on the rise [1]. Worryingly, according to Waumans et al., many of those who need mental health care do not seek it, or alternatively encounter barriers that discourage help-seeking [2]. Interestingly, in Western European countries, it appears that lack of mental health literacy, social stigma, waiting lists and logistical difficulties constitute the predominant barriers to help-seeking behaviours, as opposed to financial and/or socio-economic constraints [2,3]. There is also a need to educate citizens about mental health to minimise stigma and increase confidence in taking appropriate actions [4]. In order to meet the demand for mental health care and educate people about mental health, many research projects have focused on digital solutions, including artificial intelligence (AI) based services [1,3,5–7]. Although digital mental health tools have been in development for some time, the critical years of the covid-19 pandemic have contributed to long term challenges and changes in how services are delivered [8–11] However, it is important to note that not all apps that brand themselves as a "mental health app" have been rigorously evaluated, and may lack content that has been supported by the scientific literature or have not been created in collaboration with mental health professionals and potential end-users [12].

Recent advances in the areas of AI and natural language processing (NLP) have enabled a rapid increase in the number of mental health and wellbeing chatbots being developed. Some mental health chatbots that have been trialled include Woebot, Wysa, Shim, Vivibot and Tess [3,13–16]. Previous studies have utilised randomised controlled trials [3,14–16], lasting between 2-weeks and 8-weeks, with sample sizes ranging from 45 to 74 participants. Another study employed a quasi-experimental 'in-the-wild' study design [13] trialling the chatbot 'Wysa' for 8 weeks with 129 users. The chatbot known as 'Youper' was tested in the wild for acceptability and effectiveness with 4,517 users [17].

Researchers found that using a mental health chatbot contributed to higher engagement with the material and a higher increase in emotional awareness compared to using an informational book [3,16]. Wysa, which brands itself as an emotionally intelligent mobile chatbot app, was found to contribute to improved mental wellbeing among those who used it often [13]. While previous research on mental health chatbots shows mainly positive effects on wellbeing or higher engagement compared to traditional ways of self-management, conversations with chatbots may still feel robotic [14]. Chatbots such as Shim and Vivibot use pre-written text via a decision tree structure and both chatbots have shown mild to moderate statistically significant positive effects on mental wellbeing during trials [14,15]. While effectiveness is important, acceptability among users is equally important. Youper, for example, scored high both on user retention (over 60% more

than one week and over 40% after four weeks) and acceptability, while still showing evidence of being effective [17].

As established by previous research, the need for mental health care has increased and providers struggle to meet the current needs [1,8,9]. While it is important to increase capacity for treatment, preventative measures should also be considered. Increasing mental health literacy could help combat rising needs for mental health care by educating the public about mental health, when to seek help and reduce stigma [2,18]. Increasing mental health literacy can help improve help seeking behaviour, meaning that people may seek out help earlier, reducing the impact of the illness [2,19]. Previous research surrounding mental health chatbots show promising results, meaning that there is potential for conversational chatbots to increase mental health literacy and improve mental wellbeing among users [3,13].

The ChatPal project[20], funded by the Northern Periphery and Arctic (NPA) Programme, involves the design, development and trailing of a mental health chatbot to support the wellbeing of individuals in rural NPA areas [21]. Mental healthcare professionals were surveyed to establish their views towards health related chatbots [22] and workshops were held across NPA regions to gather user needs [5]. The ChatPal chatbot was subsequently developed based on use cases which professionals endorse and to meet the needs of end users. The chatbot is a mental health promotion tool, available in four languages including English, Scottish Gaelic, Swedish and Finnish. The chatbot is not designed to diagnose or treat those with severe mental ill health, but instead helps to promote good mental health and wellbeing. This paper focuses on the evaluation of a pre-post intervention study for the ChatPal chatbot.

The aim of this study is to establish if using a multilingual mental health and wellbeing chatbot (ChatPal) has an effect on mental wellbeing. Secondary objectives of the research are to establish the characteristics of those who had improved mental wellbeing compared to those whose mental wellbeing declined and identify themes through qualitative analysis of user feedback.

Methods

The methodology for this study was a single-arm pre-post intervention study which involved the recruitment of participants to use the ChatPal chatbot for a 12-week period. Participants' wellbeing was measured at baseline, after 6 weeks and at the end of 12-weeks and feedback was gathered through the chatbot, surveys and focus groups. This was a mixed methods study that included a quantitative evaluation of pre and post mental wellbeing scale scores, analysis of chatbot usage data and qualitative analysis of participant feedback.

Ethical Considerations

This study received ethical approval from each institution involved in the research; Ulster University Research Ethics Committee (Reference numbers REC.21.0021 and FCPSY-21-038-A), Munster Technological University Research Ethics Committee

(Reference Number MTU21034A), The Ethics Review Authority in Sweden (Reference Etikprövningsmyndigheten number 2020-00808) and The University of Eastern Finland Committee on Research Ethics gave a supporting statement (Statement 14/2021). All participants gave written informed consent to take part in the study. No form of compensation, monetary or otherwise, was offered to participants. All data collected from participants, including chatbot event log data and surveys were anonymous.

Intervention

The ChatPal chatbot was developed using Rasa (backend) and PhoneGap (frontend) with communication between backend using HTTP requests/responses (Multimedia Appendix 1). Upon receiving user inputs from the ChatPal app, the Rasa stack forwards the user input to the Rasa Natural Language Understanding (NLU) unit, which extracts user intentions (intents) and relevant metadata (entities) from the input. Once the intents and entities are identified, the chatbot response to the user is decided by the Rasa core. Rasa supports various responses including text messages, buttons, images, videos, and reminders, all of which were used in ChatPal. Rasa SDK was used to execute custom actions separately in an action server. All user interactions with ChatPal (user event log data) were stored in Amazon Relational Database Service
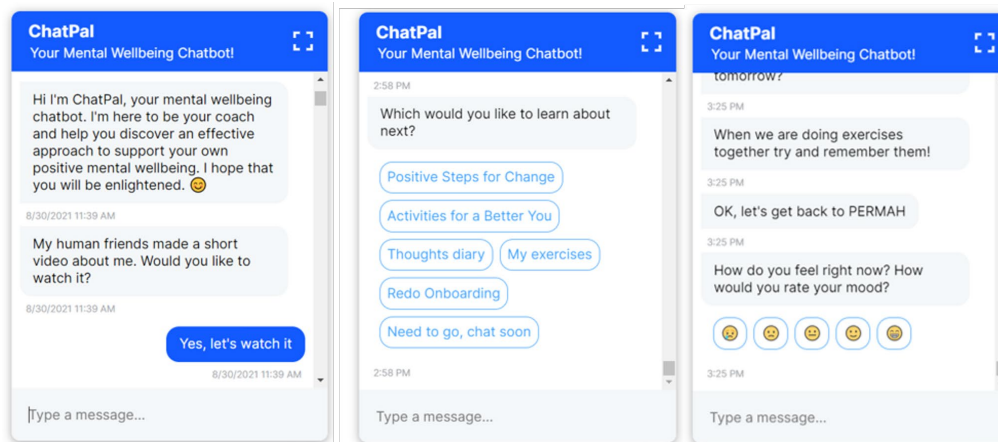
Figure 1: Sample dialogues from the ChatPal chatbot app

The overall design philosophy used in ChatPal is based on positive psychology and more specifically the PERMA/H model [23,24], which includes six core components: Positive Emotions, Engagement, Relationships, Meaning, Accomplishment and Health. ChatPal contains psychoeducational content and exercises such as mindfulness and breathing, mood logging, gratitude, and thoughts diaries among other features. All features and content available in the app can be found in multimedia appendix 2. A limited amount of data were available to train the AI in all four languages, thus within the chatbot, users can select from predefined responses (Figure 1) however ChatPal also has some AI capabilities so the chatbot will try to understand what the user is asking and direct them to the appropriate conversation.

However, if the chatbot does not understand the user input then a fallback message stating the chatbot has limited intelligence will appear and direct users to crisis helplines should they need further help. The full dialogue datasets containing the chatbot utterances in all four languages are freely available online [25].

Recruitment

Participants from rural NPA areas of Europe were recruited to use the chatbot for 12-weeks. Recruitment took place in Northern Ireland (through Ulster University), Republic of Ireland (through Munster Technological University), Scotland (through third sector organisations as coordinated by NHS Western Isles), Finland (through University of Eastern Finland) and Sweden (through Luleå University of Technology and professional networks at Norrbotten Association of Local Municipalities). Eligible participants completed online consent forms confirming they met the inclusion criteria for the study as below:

- Lived in a rural area
- Had no previous mental health diagnosis
- Had no previous suicidal thoughts and behaviours in the past year
- Were aged 18 years or older
- Consented to storage and analysis of anonymous user interactions and other chatbot data

Mental health service users were recruited through Action Mental Health in Northern Ireland. Participants filled in an online consent form confirming they met the inclusion criteria. Inclusion criteria for these participants was as follows:

- Lived in a rural area
- Were aged 18 years or older
- Had a history of mild-moderate anxiety and/or depression.
- Were not currently undergoing crisis intervention with medical services at time of study or within 12 weeks prior to the study
- Consented to storage and analysis of anonymous user interactions and other chatbot data

Study Design

Study participants were recruited online via email to take part in a 12-week single-arm pre-post intervention study with the ChatPal chatbot as detailed in the previous section. Once eligible participants completed the consent form, they were emailed further instructions on how to download ChatPal to their device and were given a region specific code to enter during onboarding. Once participants had successfully downloaded the chatbot and entered their regional code, they were automatically assigned an anonymous identifier (unique ID code for each participant) and were presented with a link to complete the baseline survey. The baseline survey asked for participants age, gender, occupation, education level, and outcome measures. The rest of the onboarding experience involved participants consenting to the storage and analysis of their anonymous chatbot usage data. The regional code and

anonymous identifier were recorded in all the surveys so demographic and outcome measures could be linked with user event log data (chatbot usage data). Thus, all data collected were anonymous and linked using participants' anonymous identifier.

Outcome measures

The primary outcome measure was the Short Warwick Edinburgh Mental Wellbeing Scale (SWEMWBS) [26] and secondary outcome measures included the Satisfaction with Life Scale (SWLS) [27] and World Health Organisation- Five Well-Being Index (WHO-5) [28]. All outcome measures were recorded at baseline, and participants received notifications to complete outcome scales again at mid-point (week 6) and at the endpoint of the study (week 12). Reminder emails were sent to participants throughout the 12-week period to encourage usage of ChatPal and completion of the surveys.

Statistical Analysis

R studio (version 3.6.0) and the R programming language were used for data wrangling and statistical analysis. Participant demographics including gender and age range were obtained from the chatbot event log data, as these questions were presented to users during onboarding. Participants' actual age and occupation were ascertained from the baseline survey. SWEMWBS scores were calculated by summing the scores for each of the seven items and transforming raw scores into metric scores using the SWEMWBS conversion table [26]. WHO-5 scores were calculated by summing the raw scores and multiplying by 4 to get a total out of 100. SWLS scores were calculated by summing the raw scores to get a total out of 35. Data were assessed for normality using the Shapiro-Wilk test, with $p < .05$ indicating data were not normally distributed. Summary statistics were computed for all outcome measures (SWEMWBS, WHO-5 and SWLS). All outcome measures followed a normal distribution so parametric tests were applied. A one-way ANOVA was used to compare scale scores from baseline, midpoint and endpoint, and t-tests were used to compare scores between groups, with $p < .05$ considered statistically significant. Chatbot event log data was filtered to only include users that enrolled on the 12-week study and summary statistics computed for each user to include; total number of interactions, number of unique days they used the chatbot and duration of use in days. The log data was then linked with survey data using the anonymous identifier present in both datasets.

Qualitative Analysis

Feedback in the form of free text responses to open ended questions were collected over the 12-week study period. Responses obtained in Swedish, and Finnish were translated to English for analysis. This feedback was then analysed using thematic analysis techniques based on those suggested by Braun and Clarke, resulting in a thematic map where the results are reported as frequencies [29].

Results

Participant demographics

A total of 348 participants took part in the intervention study. The majority of participants were female (n=254, 73%) female (Table 1). Participants ranged in age from 18-73, with a mean age of 29.6 (11.7 SD) and the most common age group was between 18-24 (n=139, 40%) (Table 1). Participants had a wide range of occupations, with the top five including students (n=105), teachers (n=16), those that were unemployed (n=11), assistant nurses (n=7), store employees (n=6).

At the beginning of the study, 81.3% (n=283) of total participants completed the baseline survey, while 11.8% (n=41) completed the mid-point survey after 6-weeks and 6.6% (n=23) completed the end-point survey at week 12.

Table 1: Demographic and mental wellbeing scale scores for participants at baseline

Age		N(%)
	18-24	139 (40)
	25-34	87 (25)
	35-44	41 (12)
	45-54	40 (11)
	55-64	37 (11)
	>65	4 (1)
Gender		N(%)
	Male	78 (22)
	Female	254 (73)
	Other gender	4 (1)
	Unknown gender	12 (4)
Outcome measures		Mean (SD)

	Short Warwick Edinburgh Mental Wellbeing Scale (SWEMWBS)	20.43 (3.66)
	The World Health Organisation- Five Well-Being Index (WHO-5)	47.67 (18.22)
	Satisfaction with Life Scale (SWLS)	20.43 (6.7)

Changes in mental wellbeing

A subset of participants (n=98) responded to a feedback survey within the chatbot. Respondents were asked how useful they thought the chatbot was for supporting mental wellbeing. The majority of participants (53%) thought the chatbot was 'somewhat useful' (n=52), while 18 (18%) thought it was 'very useful', 10 (10%) rated it as neutral (neither useful or useless), (11%) thought it was somewhat useless and the remaining 7 (7%) thought it was not useful at all. Respondents were also asked how much they believed the chatbot changed their mental wellbeing and the majority thought their wellbeing was slightly improved (42%) or unchanged (41%) as a result of using ChatPal (Figure 2). Six people reported that using the chatbot affected their mental wellbeing in a negative way, (n=3 slightly worse, n=3 significantly worse) (Figure 2). Five out of six respondents left written feedback explaining why they thought the chatbot negatively impacted their wellbeing (Textbox 1).

Textbox 1: Feedback from Swedish and Finnish participants who responded to a survey stating the chatbot made their wellbeing slightly worse or significantly worse

"Det går väldigt segt och uppfyller inte mina förväntningar varken i användarvänlighet eller vad de olika delarna handlar om." (It's very slow and doesn't meet my expectations, neither when it comes to user-friendliness nor what the different sections are about.)

"Ei ymmärtänyt mistä haluan puhua ja kun kysyin kysymyksen, alkoi tarjota hyödyttömiä ohjeita ihan randomisti." (Didn't understand what I wanted to talk about and when I asked the question, started offering useless instructions just randomly.)

"Nyt alkoi kiukustuttaa kun päästäkseni alkuvalikkoon meni kohtuuttoman kauan aikaa... Chatpal vastasi todella hitaasti, serveri ei toiminut." (Now I was beginning to feel annoyed because it took an unreasonably long time to get to the main menu... Chatpal responded very slowly, the server did not work.)

> *"Chatpalin kommentteja ei jaksa odottaa, on aivan liian hidas. Nyt tuli paljon virheitä. En saanut tehtyä haluamaani harjoitetta kohtuullisessa ajassa virheiden vuoksi."* (I'm tired of waiting for Chatpal's comments, it's way too slow. There were lots of mistakes this time. I was not able to finish the exercise I wanted within a reasonable time.)
>
> *"Chat palin käyttöön ei ole ollut nyt motivaatiota vaikka ymmärrän, että sen tehtävänä olisi juurikin auttaa jaksamaan. Ohjelman tekniset ongelmat haastavat eniten käytön aloituskynnystäni. Jos toimisi sujuvasti, olisin käyttänyt tod näk useammin ihan 'periaatteesta'* (I haven't had motivation to use Chatpal lately, even though I understand that the aim of the chatbot is to help me to carry on. Technical issues raise the most my threshold to start using it. If it worked smoothly, I probably would have used it more often as a matter of principle.)
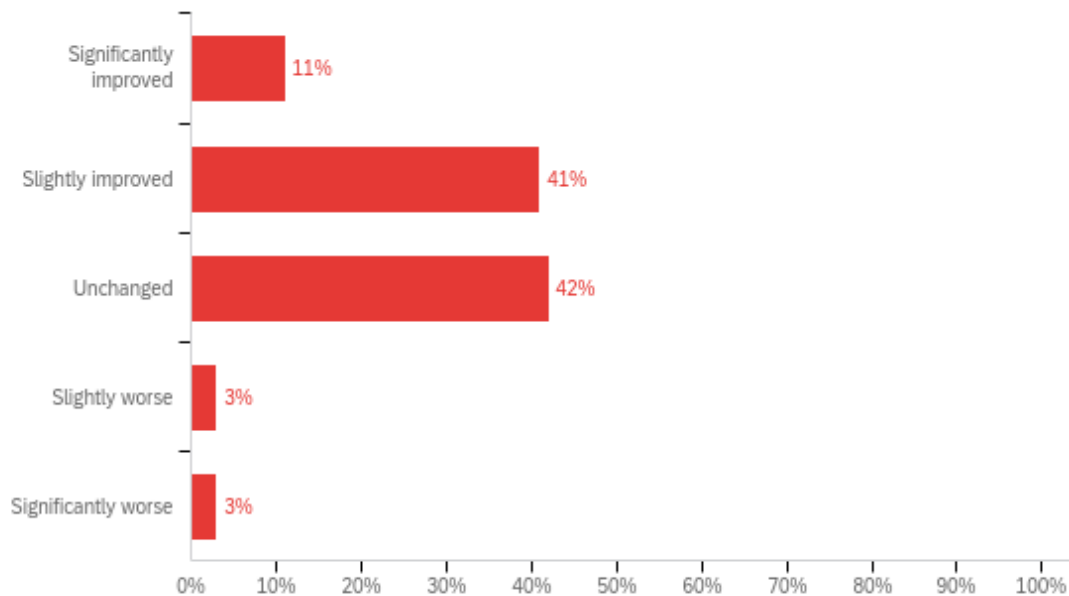
Figure 2: Participants' views on how much the chatbot changed their mental wellbeing (n=98).

The Short Warwick Edinburgh Mental Wellbeing Scale (SWEMWBS) was used to measure wellbeing throughout the study period. Wellbeing scores improved for individuals from baseline to mid-point and from baseline to end-point (Table 2, Figure 3), however the improvement in scores was not statistically significant (P=.42). Across genders, SWEMWBS scores improved non-significantly from baseline to midpoint (females: P=.99, males: P=.18) and from midpoint to end-point (females: P=.0.82, males: P=.34). SWEMWBS has a mean score of 23.5 and a standard deviation of 3.9 in the United Kingdom general population [30]. Participants in this study had an average score of 20.4 at baseline and 21.2 at the end of the ChatPal study (Table 2), indicating wellbeing scores in this population were lower on average compared to the population norms.

Secondary outcome measures included The World Health Organisation- Five Well-Being Index (WHO-5) and Satisfaction with Life Scale (SWLS). Participants' scores on the secondary outcome measures improved throughout the study period, however the improvements were not statistically significant on WHO-5 (P=.52) or SWLS (P=.81) (Figure 5, Table 3, Table 4). The WHO-5 general population norm for Finland, Sweden, United Kingdom and Ireland combined is 63.05 [31] and a cut-off score of ≤50 is typically used for depression (Topp et al. 2015). At baseline, participants' average score was 47.7 but this figure rose to 51.0 by the end of the 12-weeks (Table 3). Across genders, WHO-5 scores improved non-significantly from baseline to midpoint (females: P=.58, males: P=.41) and from midpoint to end-point (females: P=.0.90, males: P=.55). At baseline, SWLS scores averaged 20.4 (Table 4), indicating neutral satisfaction with life, and this score improved to 21.2 (Table 4) at midpoint, which indicates individuals were 'slightly satisfied' with life, according to the benchmark scoring for this scale [32].
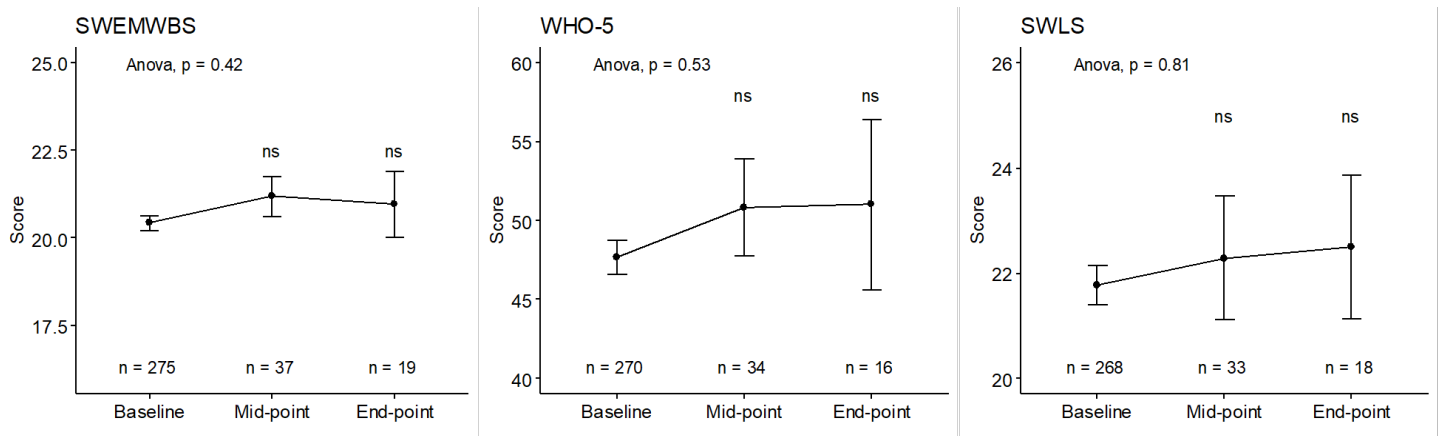
Figure 3: Short Warwick Edinburgh Mental Wellbeing scale (SWEMWBS), World Health Organisation- Five Well-Being Index (WHO-5) and Satisfaction with Life Scale (SWLS) scores across the study period. Mean and standard error for baseline, midpoint, and endpoint scores along with significance (mid and end-point scores compared to baseline).

Table 2: SWEMWBS scores for participants throughout the 12-week study period.

	Range	Mean (SD)	Median (IQR)	High wellbeing a n, (%)	Low wellbeing b n, (%)
Baseline (n=275)	9.51 - 35.00	20.43 (3.66)	19.98 (4.65)	15 (5.45)	117 (42.55)
Mid-point (n=37)	15.32 - 29.31	21.19 (3.68)	20.73 (5.52)	3 (8.11)	18 (48.64)

End-point (n=19)	12.40 - 29.31	20.95 (4.50)	21.54 (6.04)	1 (5.26)	8 (42.11)

^a Top 15% population scores (27.5-35), ^b Bottom 15% population scores (7.0-19.5)

Table 3: WHO-5 scores for participants throughout the 12-week study period.

	Range	Mean (SD)	Median (IQR)	High wellbeing[a] N, (%)	Low wellbeing[b] N, (%)
Baseline (n=270)	4-88	47.67 (18.22)	48.00 (24.00)	115 (43)	155 (57)
Mid-point (n=34)	12-84	50.82 (18.01)	48.00 (20.00)	16 (47)	18 (53)
End-point (n=16)	12-76	51.00 (21.54)	60.00 (26.00)	11 (69)	5 (31)

^a Score of ≥50, ^b Score of <50

Table 4: SWLS scores for participants throughout the 12-week study period.

	Range	Mean (SD)	Median (IQR)
Baseline (n=268)	9.51 - 35	20.43 (6.7)	19.98 (4.65)
Mid-point (n=33)	15.32 - 29.31	21.19 (3.7)	20.73 (5.52)
End-point (n=18)	12.40 - 29.31	20.95 (4.5)	21.54 (6.04)

A higher number of participants completed the primary outcome measure (SWEMWBS) at baseline and midpoint (n=37) compared to those who completed all three baseline, midpoint and end-point surveys (n=12). Of the 37 that completed the first two surveys, 16 individuals (43%) had higher SWEMWBS scores at midpoint, indicating improved wellbeing (mean change = 2.34, SD=2.24), 19 individuals (51%) had lower SWEMWBS scores at midpoint, indicative of declining mental wellbeing from baseline-midpoint (mean change=2.43, SD=1.77), and the remaining 2 (6%) had no change in scores. The group that displayed improved mental wellbeing from baseline-midpoint were significantly younger (mean age 27) compared to those who had worse mental wellbeing (mean age =37) (P=.025), however there were no significant differences in gender between the two groups (X-squared = 0.92, df = 1, P= .34). The group that had improved mental wellbeing typically had a shorter

duration of use with the chatbot, but during this time interacted with the chatbot on more unique days and logged a higher number of interactions than the group who had decreased mental wellbeing (Figure 4).
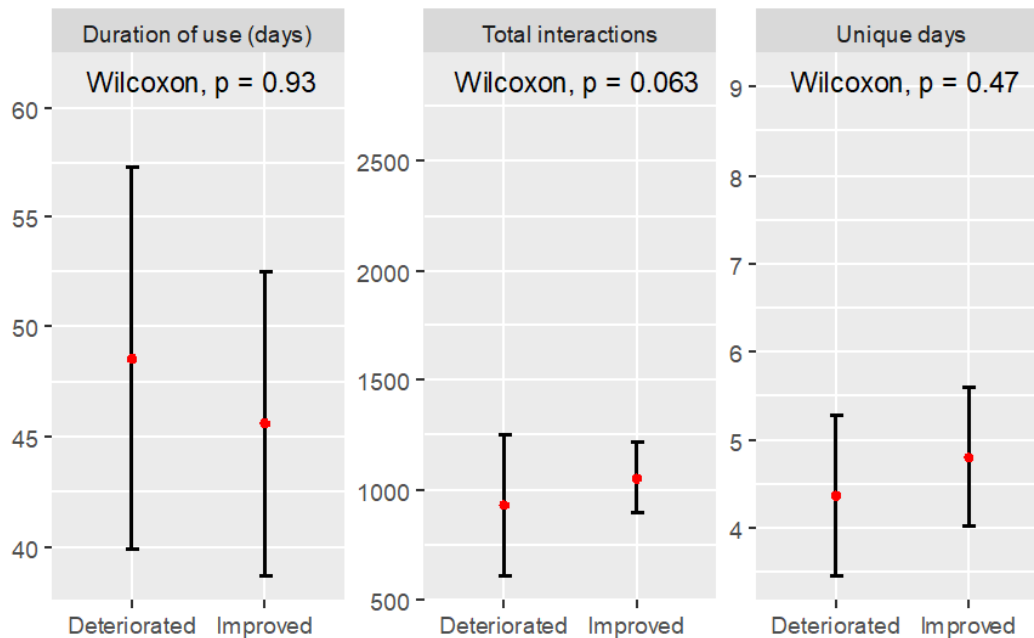
Figure 4: Usage statistics for those who improved in SWEMBS from baseline to midpoint (mean and standard error).

Qualitative analysis of user feedback

Thematic analysis was conducted on written free-text feedback (n=87) from participants. Three main themes were identified: "Positive experiences", "Mixed/ neutral experiences" and "Negative experiences" (Figure 5). Across these main themes, four common sub themes emerged including "development", "feedback", "performance" and "general comments", with an additional sub-theme "technical issues" which was specific to "negative experiences", and "trust" which was linked to "mixed/ neutral experiences".

When discussing positive aspects around feedback and general comments, participants commented that they liked the different exercises offered in the chatbot, and activities such as the gratitude diary which prompted feelings of positivity (Figure 5). Some individuals who left positive comments gave suggestions for future development, for example a place to store favourite exercises and being able to set reminders to use the app (Figure 5).

Some participants had a mixed or overall neutral experience with the chatbot, with one person stating that they liked it but still didn't want to use it, and another commenting that they liked the content, but still wanted to discuss their feelings with a human (Figure 5). Most of the comments with mixed/ neutral aspects

mentioned that they liked the chatbot overall, but that there were some barriers, such as technical errors or confusion around navigating content which needed to be overcome. Two participants left specific comments stating they would like a clearer, easy to use menu, and several other participants noted menu navigation as something that could be improved (Figure 5). Trust was mentioned, and participants stated that they would be more likely to trust the chatbot if it was transparent about its purpose and what will happen to user data at the very start (Figure 5).

A few neutral and negative experiences were noted specific to the Scottish Gaelic version of the app. Firstly, native Scottish Gaelic speakers who took part were not accustomed to seeing the language written down, as typically the language is spoken more often than it is read. Secondly, for users reading Scottish Gaelic translations, some of the language around mental health and wellbeing was not common terminology in day-to-day conversations and several participants thought that having a Gaelic glossary would be helpful.

Negative experiences mostly included technical issues with the chatbots functionality (Figure 5). At the beginning of the study, participants commented on the app performance, stating that the app was slow to respond. This information was shared with the developers so that improvements could be made while the study was ongoing, and later feedback from participants mentioned this in a positive manner, noting the chatbot responses were considerably faster (Figure 5). A few participants disliked the choice of 5 different icons (emojis) to rate their mood from happy to sad, as they thought this range wasn't wide enough to express how they felt at the time (Figure 5).
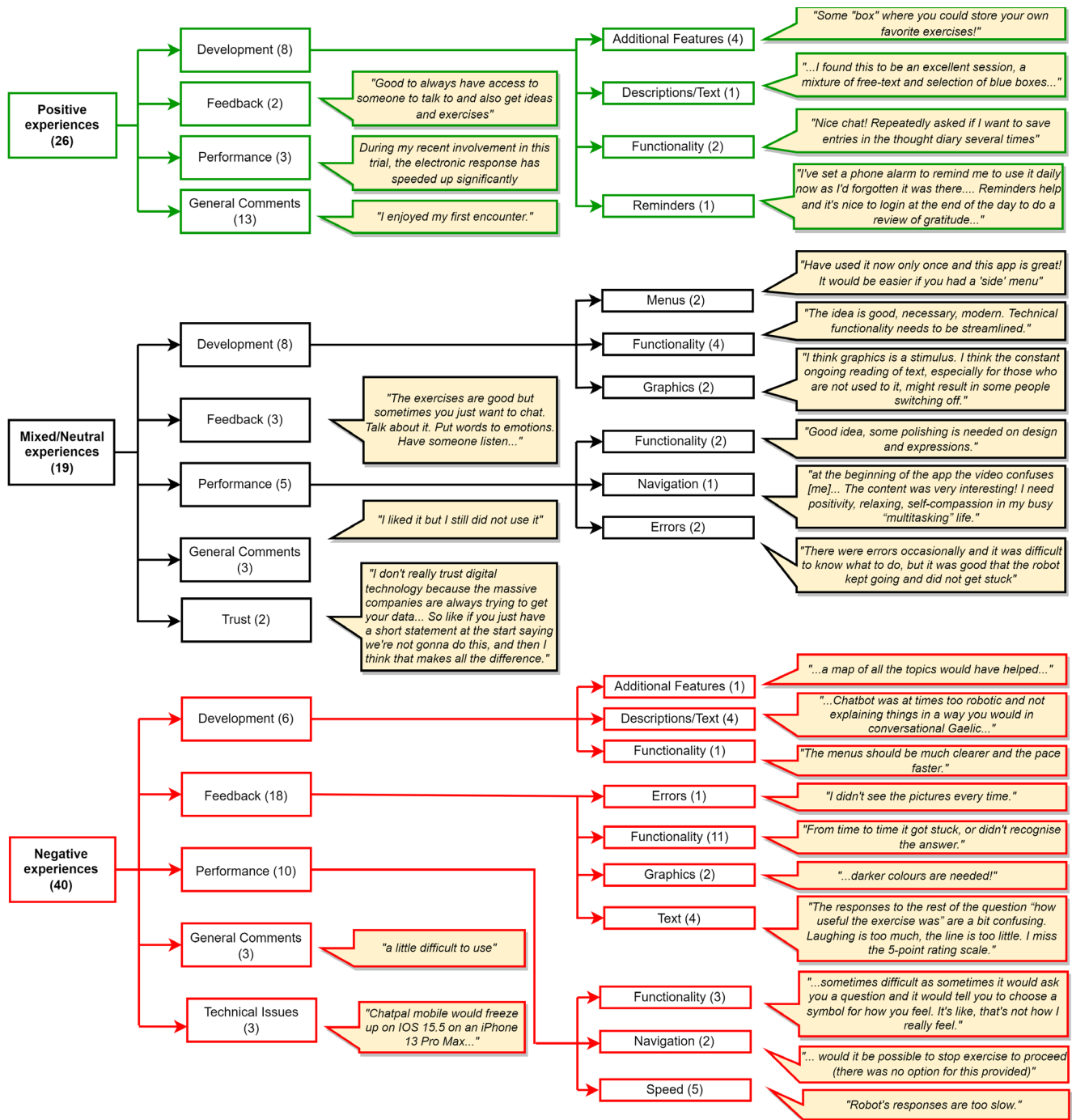
Positive experiences (26)
- Development (8)
- Feedback (2) — "Good to always have access to someone to talk to and also get ideas and exercises"
- Performance (3) — "During my recent involvement in this trial, the electronic response has speeded up significantly"
- General Comments (13) — "I enjoyed my first encounter."

Development (8):
- Additional Features (4) — "Some "box" where you could store your own favorite exercises!"
- Descriptions/Text (1) — "...I found this to be an excellent session, a mixture of free-text and selection of blue boxes..."
- Functionality (2) — "Nice chat! Repeatedly asked if I want to save entries in the thought diary several times"
- Reminders (1) — "I've set a phone alarm to remind me to use it daily now as I'd forgotten it was there.... Reminders help and it's nice to login at the end of the day to do a review of gratitude..."

Mixed/Neutral experiences (19)
- Development (8)
- Feedback (3) — "The exercises are good but sometimes you just want to chat. Talk about it. Put words to emotions. Have someone listen..."
- Performance (5)
- General Comments (3) — "I liked it but I still did not use it"
- Trust (2) — "I don't really trust digital technology because the massive companies are always trying to get your data... So like if you just have a short statement at the start saying we're not gonna do this, and then I think that makes all the difference."

Development (8):
- Menus (2) — "Have used it now only once and this app is great! It would be easier if you had a 'side' menu"
- Functionality (4) — "The idea is good, necessary, modern. Technical functionality needs to be streamlined."
- Graphics (2) — "I think graphics is a stimulus. I think the constant ongoing reading of text, especially for those who are not used to it, might result in some people switching off."

Performance (5):
- Functionality (2) — "Good idea, some polishing is needed on design and expressions."
- Navigation (1) — "at the beginning of the app the video confuses [me]... The content was very interesting! I need positivity, relaxing, self-compassion in my busy "multitasking" life."
- Errors (2) — "There were errors occasionally and it was difficult to know what to do, but it was good that the robot kept going and did not get stuck"

Negative experiences (40)
- Development (6)
- Feedback (18)
- Performance (10)
- General Comments (3) — "a little difficult to use"
- Technical Issues (3) — "Chatpal mobile would freeze up on IOS 15.5 on an iPhone 13 Pro Max..."

Development (6):
- Additional Features (1) — "...a map of all the topics would have helped..."
- Descriptions/Text (4) — "...Chatbot was at times too robotic and not explaining things in a way you would in conversational Gaelic..."
- Functionality (1) — "The menus should be much clearer and the pace faster."

Feedback (18):
- Errors (1) — "I didn't see the pictures every time."
- Functionality (11) — "From time to time it got stuck, or didn't recognise the answer."
- Graphics (2) — "...darker colours are needed!"
- Text (4) — "The responses to the rest of the question "how useful the exercise was" are a bit confusing. Laughing is too much, the line is too little. I miss the 5-point rating scale."

Performance (10):
- Functionality (3) — "...sometimes difficult as sometimes it would ask you a question and it would tell you to choose a symbol for how you feel. It's like, that's not how I really feel."
- Navigation (2) — "... would it be possible to stop exercise to proceed (there was no option for this provided)"
- Speed (5) — "Robot's responses are too slow."

Figure 5: Themes identified from ChatPal feedback, with examples of user comments for each subtheme.

Discussion

Principal Results

This study reports on the evaluation of a multilingual chatbot named ChatPal, designed as a mental health promotion tool to promote positive mental wellbeing in individuals across sparsely populated areas of Europe. A total of 348 participants were recruited to use the chatbot for a 12-week period, with outcome measures recorded at the beginning, after 6 weeks (midpoint) and after 12 weeks (endpoint). This study benefited from the inclusion of participants with a wide range of demographics, including adults from the ages of 18 to 65+. In response to the survey question which asked participants how they believed the chatbot improved their mental wellbeing, over half (52%) said that it significantly or slightly improved. Only a few participants (n=6) thought the chatbot negatively impacted their wellbeing, and the reasons for this were due to technical errors or frustration at the slow speed of chatbot responses, which was later rectified. This is an important point for trials of digital interventions to address i.e. that any technical issues can be an unwanted confounder and negatively affect the outcome variables. Hence, it is crucial that technical testing is rigorous and that studies incorporate a pilot to address technical issues or barriers ahead of the main intervention study or trial. The outcome measures which were recorded at baseline, midpoint and endpoint revealed that participants' wellbeing generally improved throughout the study period, however the improvements in wellbeing were not statistically significant for SWEMWBS, WHO-5 and SWLS. Non-significant improvements in wellbeing were similar for both males and females. The SWEMWBS scores at baseline and throughout the 12-weeks remained below the population norms for the UK [30]. However, by the end of the trial WHO-5 scores improved to 51, which was just above the threshold for depression (score of 50 or below) compared to the scores at baseline which were below this threshold.

It is important to characterise users that showed improvements in wellbeing after using ChatPal, as even if it only helps a small proportion of users then it could be seen as beneficial. Individuals that did show improvements in wellbeing (n=16) interacted more with the chatbot and were significantly younger on average (27) compared to those whose wellbeing worsened over the study, with an average age of 37. This may be related to the likely acceptance and adoption of text messaging and chatbots amongst younger generations who are also considered to be the 'generation mute'. Generation mute is the nickname given to young people aged under 25, who have grown up in the digital era and predominantly use smartphones for messaging as opposed to voice calls. It may be the case that those who are younger and are more familiar with text based applications benefit more than those who are older.

Given that the results showed that marginal improvements in mental wellbeing were seen in those who used ChatPal, the app could be thought of as just one tool of

many in a larger service offering, that could be used in conjunction with other digital or face-to-face services. There is potential for ChatPal to be used with those who are on waiting lists for mental health services, or outpatients who have already received counselling for example. ChatPal could be thought of as a health promotion tool that can be used by anyone in the general population, providing means of learning and education about what it means to be in a good state of mental wellbeing and as a preventative tool to stop mental health problems escalating. It can also provide information about mental health, signs of mental illness and tips about giving support which can improve mental health literacy in the population and increase public acceptance of mental health care [33]. Educating people about mental health may help to minimise stigma and increase confidence in taking appropriate action [4]. Participants stated they were somewhat unfamiliar with mental health terminology in Scottish Gaelic, this further supports the use of ChatPal or other mental health promotion apps as educational tools that could be used to promote wellbeing in different languages.

Some participants mentioned that they would prefer to open up to a human regarding their feelings. Others may be more likely to open up to a chatbot as it allows them to remain anonymous. Previous studies have shown patients disclosed more sensitive information to a chatbot when compared to a human therapist [34] and that people sometimes engage more on a personal level with virtual therapists [16].

Ultimately, a personalised approach for mental health and wellbeing would be beneficial, as what works best for each person needs to be considered on an individual basis and some may prefer to use multiple different tools to look after their mental wellbeing.

Comparison with Prior Work

This paper further supports the need for blended service offerings, given that mental wellbeing scores didn't significantly increase. This result is in line with findings from recent reviews [7,35] which found that blended service offerings were more effective in treating those living with depression, and several studies have shown that guided interventions resulted in better adherence and lower dropouts compared with unguided interventions [36–38].

Previous studies [14,15] utilised chatbots with a decision tree structure and prewritten text, which is similar to the ChatPal chatbot. Almost all other mental health chatbots, for example Woebot [3] and Wysa [13] are targeted at treating mental ill health using cognitive-behavioural therapy and dialectical behaviour therapy, unlike ChatPal which focuses more on mental health promotion. ChatPal is built around the PERMA/H positive psychology framework [23,24], and includes evidence-based practices. The ChatPal chatbot is also available in four languages including English, Swedish, Finnish and Scottish Gaelic which is unique. Previously, AI mental health chatbots have occasionally generated inappropriate responses [39], which can be problematic if the user is disclosing concerning information

about their mental health. ChatPal contains limited AI and if it does not understand user input then it will signpost the user to crisis helplines should the user need additional help.

A recent paper by Torous et al. highlighted the future importance of accelerated use of digital mental health tools, but also pointed out that there are barriers to overcome such as training staff, high quality evidence and digital equity [40]. In this study, only a small subset of users who started the study actually completed it. Adherence is a challenge with digital health intervention studies in general, and high drop off is typical with mental health apps. A previous study looked at real-world data of health app usage and found that on average only 3.9% users continued to use these apps after 15 days [41]. In future, it will be important to find strategies that will increase engagement and adherence in mental health app trials.

Limitations

This study included a wide age range of participants, however the majority of participants were female (75%). The chatbot study took place across four European countries, but a high proportion of participants were from Sweden. As the study design was a single-arm pre-post study, another limitation is lack of a randomised controlled environment and control group which has the potential to cause bias in the study and limits the findings regarding the chatbot's causality on mental wellbeing. In addition, no power calculation was conducted a priori. Given the small number of participants that completed the outcome measures at endpoint (n=19) and that we didn't find any significant changes in wellbeing scores, this could be the result of small sample size and underpowering. The investigators did not select specific individuals for inclusion in the study, however as participants self-selected to take part in the study based on the inclusion criteria, it is possible that self-selection bias may have been introduced into the study. Participant adherence in the study may have been better if investigators were able to interact with participants face-to-face, however most interactions and engagement was carried out online due to the pandemic. However, this study could be considered as somewhat 'real world' given the limited interaction participants had with researchers (e.g. consent was taken electronically). The participant feedback obtained was anonymous and could not be linked to the survey data, so we were unable to look at specific feedback across age groups, gender and those who exhibited changes in wellbeing.

Conclusions

The ChatPal chatbot had a marginal effect on mental wellbeing although this was not statistically significant. Interestingly, the majority of users believed that the chatbot slightly or significantly improved their mental wellbeing. Those who had improved mental wellbeing had more interactions with the chatbot and were younger in age. Future work could also include analysing feedback from chatbot users for specific age groups, as requirements for mental health chatbots and digital technologies in general should be considered across different generations. The ChatPal chatbot can be seen as a mental health promotion tool that targets the general population rather than treating those with mental ill health. The results of this study support this, and

we propose that ChatPal could be included as part of a blended service that complements other digital tools as well as face-to-face service offerings. Further research is needed to confirm the efficacy of ChatPal alongside other complementary services.

Data Availability
The ChatPal chatbot dialogue data which contains the chatbot utterances in English, Swedish, Finnish and Scottish Gaelic are freely available from Ulster University and can be accessed using this link https://doi.org/10.21251/eefafafa-705c-4cdb-921e-a2bb0d32b6ef. The survey and event log datasets data sets generated during and/or analysed during the current study are not publicly available as participants consent included data sharing within the project research team only.

Conflicts of Interest
None declared.

Abbreviations
AI: Artificial intelligence
NPA: Northern Periphery and Arctic
SWEMWBS: Short Warwick Edinburgh Mental Wellbeing Scale
SWLS: Satisfaction with Life Scale
WHO-5: World Health Organisation- Five Well-Being Index

References

1. Timakum T, Xie Q, Song M. Analysis of E-mental health research: mapping the relationship between information technology and mental healthcare. BMC Psychiatry. BioMed Central Ltd; 2022;22:1–17. PMID: 35078432

2. Waumans RC, Muntingh ADT, Draisma S, Huijbregts KM, van Balkom AJLM, Batelaan NM. Barriers and facilitators for treatment-seeking in adults with a depressive or anxiety disorder in a Western-European health care setting: a qualitative study. BMC Psychiatry. BioMed Central Ltd; 2022;22:1–15. PMID: 35247997

3. Fitzpatrick KK, Darcy A, Vierhile M. 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 Ment Heal. 2017;4(2):e19.

4. Kostenius C, Gabrielsson S, Lindgren E. Promoting Mental Health in School— Young People from Scotland and Sweden Sharing Their Perspectives. Int J Ment Health Addict. Springer; 2020;18(6):1521–1535.

5. Potts C, Ennis · E, Bond · R B, Mulvenna · M D, Mctear · M F, Boyd · K, et al. Chatbots to Support Mental Wellbeing of People Living in Rural Areas: Can User Groups Contribute to Co-design? J Technol Behav Sci 2021 64. Springer; 2021;6(4):652–665.

6. Zidaru T, Morrow EM, Stockley R. Ensuring patient and public involvement in the transition to AI-assisted mental health care: A systematic scoping review and agenda for design justice. Heal Expect. John Wiley & Sons, Ltd; 2021;24(4):1072–1124. PMID: 34118185

7. Moshe I, Terhorst Y, Philippi P, Domhardt M, Cuijpers P, Cristea I, et al. Digital Interventions for the Treatment of Depression: A Meta-Analytic Review. Psychol Bull. 2021. PMID: 34898233

8. Fiorillo A, Gorwood P. The consequences of the COVID-19 pandemic on mental health and implications for clinical practice. Eur Psychiatry. Cambridge University Press; 2020;63(1):e32. PMID: 32234102

9. Nowrouzi-Kia B, Stier J, Ayyoub L, Hutchinson L, Laframboise J, Mihailidis A. The Characteristics of Canadian University Students' Mental Health, Engagement in Activities and Use of Smartphones: A descriptive pilot study. Heal Psychol Open. SAGE PublicationsSage UK: London, England; 2021;8(2):1–8.

10. Satre DD, Meacham MC, Asarnow LD, Fisher WS, Fortuna LR, Iturralde E. Opportunities to Integrate Mobile App-Based Interventions Into Mental Health and Substance Use Disorder Treatment Services in the Wake of COVID-19. Am J Health Promot. Am J Health Promot; 2021;35(8):1178–1183. PMID: 34652971

11. Piper S, Davenport TA, LaMonica H, Ottavio A, Iorfino F, Cheng VWS, et al. Implementing a digital health model of care in Australian youth mental health services: protocol for impact evaluation. BMC Health Serv Res. BioMed Central Ltd; 2021;21(1):1–9. PMID: 33980229

12. Michel T, Tachtler F, Slovak P, Fitzpatrick G. A review of youth mental health

promotion apps towards their fit with youth media preferences. EAI Endorsed Trans Pervasive Heal Technol. European Alliance for Innovation; 2019;5(17).

13. Inkster B, Sarda S, Subramanian V. An Empathy-Driven, Conversational Artificial Intelligence Agent (Wysa) for Digital Mental Well-Being: Real-World Data Evaluation Mixed-Methods Study. JMIR mHealth uHealth. 59 WINNERS CIRCLE, TORONTO, ON M4L 3Y7, CANADA: JMIR PUBLICATIONS, INC; 2018;6(11).

14. Ly KH, Ly AM, Andersson G. A fully automated conversational agent for promoting mental well-being: A pilot RCT using mixed methods. Internet Interv. Elsevier B.V.; 2017;10:39–46.

15. Greer S, Ramo D, Chang YJ, Fu M, Moskowitz J, Haritatos J. Use of the chatbot "vivibot" to deliver positive psychology skills and promote well-being among young people after cancer treatment: Randomized controlled feasibility trial. JMIR mHealth uHealth. 2019;7(10):1–13. PMID: 31674920

16. Fulmer R, Joerin A, Gentile B, Lakerink L, Rauws M. Using Psychological Artificial Intelligence (Tess) to Relieve Symptoms of Depression and Anxiety: Randomized Controlled Trial. JMIR Ment Heal. 2018;5(4):e64.

17. Mehta A, Niles AN, Vargas JH, Marafon T, Couto DD, Gross JJ. Acceptability and Effectiveness of Artificial Intelligence Therapy for Anxiety and Depression (Youper): Longitudinal Observational Study. J Med Internet Res . Journal of Medical Internet Research; 2021;23(6):e26771. PMID: 34155984

18. Onnela A, Hurtig T, Ebeling H. School professionals committed to student well-being. Int J Circumpolar Health. Taylor & Francis; 2021;80(1). PMID: 33491598

19. Wei Y, McGrath PJ, Hayden J, Kutcher S. Mental health literacy measures evaluating knowledge, attitudes and help-seeking: a scoping review. BMC Psychiatry 2015 151. BioMed Central; 2015;15(1):1–20. PMID: 26576680

20. Consortium C project. ChatPal website [Internet]. 2022.

21. Northern Periphery and Arctic Programme. Programme Area [Internet].

22. Sweeney C, Potts C, Ennis E, Bond R, Mulvenna MD, O'Neill S, et al. Can Chatbots Help Support a Person's Mental Health? Perceptions and Views from Mental Healthcare Professionals and Experts. ACM Trans Comput Healthc. 2021;2(3):1–16.

23. Seligman MEP. Flourish: A visionary new understanding of happiness and well-being. New York, NY: Simon and Schuster; 2011.

24. McQuaid M, Kern P. Your wellbeing blueprint : feeling good and doing well at work. 1st ed. Michelle McQuaid Pty Ltd; 2017.

25. Bond R, Booth F, Mulvenna M, Potts C, O'Keeffe R, Kostenius C, et al. Chatpal Chatbot dialogue data - version 1 [Internet]. 2022.

26. Stewart-Brown S, Tennant A, Tennant R, Platt S, Parkinson J, Weich S. Internal construct validity of the Warwick-Edinburgh Mental Well-Being Scale (WEMWBS): A Rasch analysis using data from the Scottish Health Education Population Survey. Health Qual Life Outcomes. BioMed Central; 2009;7(1):1–8. PMID: 19228398

27. Diener E, Emmons RA, Larsem RJ, Griffin S. The Satisfaction With Life Scale. https://doi.org/101207/s15327752jpa4901_13. Lawrence Erlbaum

Associates, Inc. ; 2010;49(1):71–75. PMID: 16367493

28. Topp CW, Østergaard SD, Søndergaard S, Bech P. The WHO-5 well-being index: A systematic review of the literature. Psychother Psychosom. S. Karger AG; 2015;84(3):167–176. PMID: 25831962

29. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol. 2006;3(2):77–101.

30. Fat LN, Scholes S, Boniface S, Mindell • Jennifer, Stewart-Brown S. Evaluating and establishing national norms for mental wellbeing using the short Warwick-Edinburgh Mental Well-being Scale (SWEMWBS): findings from the Health Survey for England.

31. Eurofound. European Quality of Life Survey 2012 [Internet]. 2014.

32. Pavot W, Diener E. Happiness Experienced: The Science of Subjective Well-being. Oxford Handb Happiness. Oxford Academic; 2013.

33. Jorm AF. Mental health literacy. Public knowledge and beliefs about mental disorders. Br J Psychiatry. Br J Psychiatry; 2000;177(NOV.):396–401. PMID: 11059991

34. Lucas GM, Rizzo A, Gratch J, Scherer S, Stratou G, Boberg J, et al. Reporting mental health symptoms: Breaking down barriers to care with virtual human interviewers. Front Robot AI. Frontiers Media S.A.; 2017;451:1–9.

35. Kambeitz-Ilankovic L, Rzayeva U, Völkel L, Wenzel J, Weiske J, Jessen F, et al. A systematic review of digital and face-to-face cognitive behavioral therapy for depression. npj Digit Med 2022 51. Nature Publishing Group; 2022;5(1):1–8.

36. Baumeister H, Reichler L, Munzinger M, Lin J. The impact of guidance on Internet-based mental health interventions — A systematic review. Internet Interv. Elsevier; 2014;1(4):205–215.

37. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res. J Med Internet Res; 2011;13(3). PMID: 21821503

38. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev. Clin Psychol Rev; 2012;32(4):329–342. PMID: 22466510

39. De Nieva JO, Joaquin JA, Tan CB, Marc Te RK, Ong E. Investigating Students Use of a Mental Health Chatbot to Alleviate Academic Stress. CHIuXiD '20 6th Int ACM In-Cooperation HCI UX Conf. Association for Computing Machinery; 2020. p. 1–10.

40. Torous J, Myrick KJ, Rauseo-Ricupero N, Firth J. Digital Mental Health and COVID-19: Using Technology Today to Accelerate the Curve on Access and Quality Tomorrow. JMIR Ment Heal. JMIR Mental Health; 2020;7(3):e18848.

41. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: Systematic search and panel-based usage analysis. J Med Internet Res. JMIR Publications Inc.; 2019;21(9):e14567. PMID: 31573916