Taylor & Francis
Taylor & Francis Group

International Journal of Human–Computer Interaction

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hihc20

Investigating the Key Success Factors of Chatbot-Based Positive Psychology Intervention with Retrieval- and Generative Pre-Trained Transformer (GPT)-Based Chatbots

Ivan Liu, Fangyuan Liu, Yuting Xiao, Yajia Huang, Shuming Wu & Shiguang Ni

To cite this article: Ivan Liu, Fangyuan Liu, Yuting Xiao, Yajia Huang, Shuming Wu & Shiguang Ni (08 Jan 2024): Investigating the Key Success Factors of Chatbot-Based Positive Psychology Intervention with Retrieval- and Generative Pre-Trained Transformer (GPT)-Based Chatbots, International Journal of Human–Computer Interaction, DOI: 10.1080/10447318.2023.2300015

To link to this article: https://doi.org/10.1080/10447318.2023.2300015

Published online: 08 Jan 2024.

INTERNATIONAL JOURNAL OF HUMAN–COMPUTER INTERACTION
https://doi.org/10.1080/10447318.2023.2300015

Taylor & Francis
Taylor & Francis Group

RESEARCH REPORT

Investigating the Key Success Factors of Chatbot-Based Positive Psychology Intervention with Retrieval- and Generative Pre-Trained Transformer (GPT)-Based Chatbots

Ivan Liu[a,b], Fangyuan Liu[a], Yuting Xiao[b], Yajia Huang[b], Shuming Wu[b], and Shiguang Ni[c]

[a]Department of Psychology, Faculty of Arts and Science, Beijing Normal University at Zhuhai, China; [b]Faculty of Psychology, Beijing Normal University, China; [c]Shenzhen International Graduate School, Tsinghua University, China

ABSTRACT
Technologically assisted methods have been extensively utilized in positive psychology interventions (PPIs), which aim to elevate the happiness levels of the widespread and diverse general public, a population that traditional methodologies struggle to comprehend and impact. Nevertheless, the literature provides insufficient insights into the effectiveness of chatbot-based PPIs (Chat-PPIs). This study endeavors to fill this void by employing both retrieval-based and generative pre-trained transformer (GPT)-based chatbots and executing three randomized controlled trials involving 326 participants to investigate the hypothesized effectiveness mechanisms. The statistical analysis affirms the effectiveness of Chat-PPI. Moreover, our results indicate that personalized PPI recommendations, adaptive multi-round dialogues, and real-time feedback significantly augment the efficacy of Chat-PPI. Besides exploring and confirming the mechanisms behind Chat-PPI, this study also endorses the use of generative chatbots, which, although less controllable, provide more natural interaction that can boost the efficacy of Chat-PPI.

KEYWORDS
Chatbot; positive psychology intervention; ChatGPT; retrieval-based chatbot; GPT-based chatbot; generative chatbot; prompt engineering

1. Introduction

Pursuing happiness, its essence, and methods of attainment have been subjects of centuries-long debate across research, religion, and philosophy. In the past two decades, positive psychology has also given abundant empirical evidence on happiness and its composition (Lomas et al., 2021). Happiness serves as not only an ultimate goal but also a conduit to achieving other objectives. Sustained positive emotions widen an individual's thought-action repertoire, nurturing long-lasting personal resources beneficial for learning, development, and growth (Fredrickson, 2001). Positive emotions are also vital as psychological assets in preventing mental health disorders and can improve conditions for those already suffering (Layous et al., 2011). Moreover, mental well-being positively influences physical health (Park et al., 2014), with happier individuals demonstrating superior cardiovascular functions, more robust immune systems (Boehm & Kubzansky, 2012), and lower disease and mortality risks.

Given the importance of happiness, researchers have developed and empirically validated numerous positive psychological interventions (PPIs) aimed at bolstering positive emotions, cognitions, and behaviors (Parks & Biswas-Diener, 2013). Among these strategies, interventions designed to enhance positive cognitions, including the Three Good Things (TGT) (Seligman et al., 2005), Best Possible Self (BPS) (Layous et al., 2013), and gratitude practices (Emmons & McCullough, 2003) are prevalently integrated into various self-help psychological applications for their simplicity and practicality for implementation.

However, translating laboratory advances into real-world applications poses multiple challenges. Positive psychology studies aimed at helping the majority of people require interventions to be accessible, affordable, on-demand, and adap– to diverse and evolving environments (Waterman, 2013). Furthermore, since most PPI programs are designed as self-help methods for individuals not diagnosed with mental disorders (Howells et al., 2016), these self-help psychological interventions must be user-friendly, understandable, and resistant to user errors - enabling individuals to engage with them without needing professional assistance or formal training. Also, the effectiveness of these interventions heavily relies on participants' continuous commitment, which is frequently hindered by a lack of motivation and human support, leading to high dropout rates, especially for programs with inflexible and repeating content (Richards & Richardson, 2012).

To surmount the challenges mentioned above, psychology researchers have leveraged information technology to deliver PPIs (Schueller et al., 2013). However, most of these solutions have yet to overcome the stated issues entirely. While some technologies have made psychological interventions more accessible, efforts to reduce high dropout rates have

CONTACT Shiguang Ni ni.shiguang@sz.tsinghua.edu.cn Tsinghua Shenzhen International Graduate School, University Town of Shenzhen, Nanshan District, Shenzhen 518055, China

only partially succeeded. As a result, pioneering researchers are exploring alternatives, with chatbot-based PPIs (Chat-PPIs) showing promise.

Chatbots, or conversational agents, are computer systems developed based on natural language processing (NLP) that allow humans to interact with computers using human language (Boucher et al., 2021; Siddique & Chow, 2021). They have proven effective in a wide range of fields, such as marketing (Jain et al., 2023; Paul et al., 2023), customer service (Nicolescu & Tudorache, 2022), language learning (Haristiani, 2019), healthcare (Ayers et al., 2023; Chow et al., 2023) and medicine (Lee et al., 2023). Furthermore, ample evidence supports the utility of chatbots in psychological interventions, specifically in addressing issues like anxiety (Greer et al., 2019; Lim et al., 2022), depression (Fitzpatrick et al., 2017), stress (Medeiros et al., 2019), acrophobia (Abd-Alrazaq et al., 2020) and loneliness (Ta et al., 2020).

Compared to other technological solutions, chatbots offer several unique features. They harness natural language, the primary mode of human communication, which lessens user resistance. Unlike traditional software interfaces that require users to grasp the system design before using it, chatbots actively discern user intent, easing the learning process. This adaptability is especially advantageous for those less tech-savvy. Furthermore, chatbots can seamlessly integrate with existing voice or phone-based services. As highlighted by researchers, compared with text-based chatbots, voice assistants (voice-based chatbots) can enhance perceived efficiency and enjoyment while demanding less cognitive effort from users (Rzepka et al., 2022). Such chatbots are particularly useful when continuous screen engagement is impractical, like during driving or exercising. During interactions, chatbots can identify and comprehend user needs and emotional states (Ahmad et al., 2022), delivering adaptive, real-time feedback (Grové, 2021). Advances in AI chatbot technology, rooted in NLP, make interactions closely resemble genuine human conversations (Lee et al., 2023). This dynamic engagement ensures precise responses for users and minimizes monotony during extended sessions, a trait crucial for psychological health services (Chaves & Gerosa, 2021).

Despite the apparent potential of Chat-PPI, our preliminary literature review identified fewer than twenty studies on the topic, and a large portion of these studies did not report statistically significant results (Greer et al., 2019; Loveys et al., 2021; Ly et al., 2017). Given the widespread publication bias in research (Masicampo & Lalande, 2012), the lack of compelling evidence supporting Chat-PPI's effectiveness becomes even more striking. Moreover, most studies on Chat-PPI primarily emphasize verifying the overall effectiveness of chatbots in psychological interventions, with little exploration of the underlying mechanisms contributing to their effectiveness (Jeong et al., 2023).

This study seeks to bridge this research gap by experimentally validating chatbot effectiveness and identifying potential mechanisms that boost Chat-PPI efficacy. Previous research has highlighted that dialogue and interaction design are critical factors in chatbot effectiveness and usability (Fitzpatrick et al., 2017; Liu et al., 2022). This study will ascertain whether the user-chatbot interaction is the key to Chat-PPI's success. To this end, we will examine three specific hypotheses:

Researchers have argued that personalization is an essential feature of chatbot interaction and is crucial to enhancing the effectiveness of chatbot-based psychological intervention (Heintzelman et al., 2023; Jeong et al., 2023). Therefore, Chat-PPI should be tailored to match the user's status and intent, but existing studies neglect this possibility. Consequently, we present our first hypothesis, which we will test in sub-study 1:

H1: Chatbots delivering personalized PPI suggestions can enhance the effectiveness of the interventions.

Unlike traditional software communication designs, which largely follow a document metaphor and often present content as an extended article – as seen in most online PPIs, chatbots interact like human conversation. They guide users through multi-round dialogues, adapting their responses based on user inputs. This study proposes that this unique interaction style might be why chatbots outperform traditional online PPIs. As such, we present our second hypothesis:

H2: The adaptive multi-round interaction of chatbots can enhance the effectiveness of Chat-PPIs.

One distinctive feature that differentiates human-guided PPIs from self-help PPIs is the therapist's capacity to offer empathetic and encouraging responses (Stiles et al., 1998). However, given the advances in artificial intelligence, chatbots are now capable of delivering proper feedback based on user inputs in real-time (Cheng et al., 2023). As a result, we posit our third hypothesis:

H3: The ability of chatbots to provide empathetic and encouraging real-time feedback enhances the effectiveness of Chat-PPIs.

2. Methods

2.1. Sub-study 1

In sub-study 1, we recruited 256 online participants from China, of whom 207 completed the questionnaire, and 154 (68.2% female, aged 18–55) fulfilled all the experimental requirements (Figure 1). The Ethics Committee of the first author's institution granted ethical approval, and all participants provided written informed consent. Each participant received RMB 25 (approximately US$ 3.7) as compensation for their participation.

Upon completion of the questionnaire, this study divided participants into three groups. In the recommendation group, we employed a chatbot to determine the most appropriate PPI for the participants. The TGT exercise, which involves sharing positive experiences, was recommended to participants in a relatively upbeat mood. Conversely, the BPS intervention, which aims at bolstering optimism, was suggested for participants manifesting less positive emotional

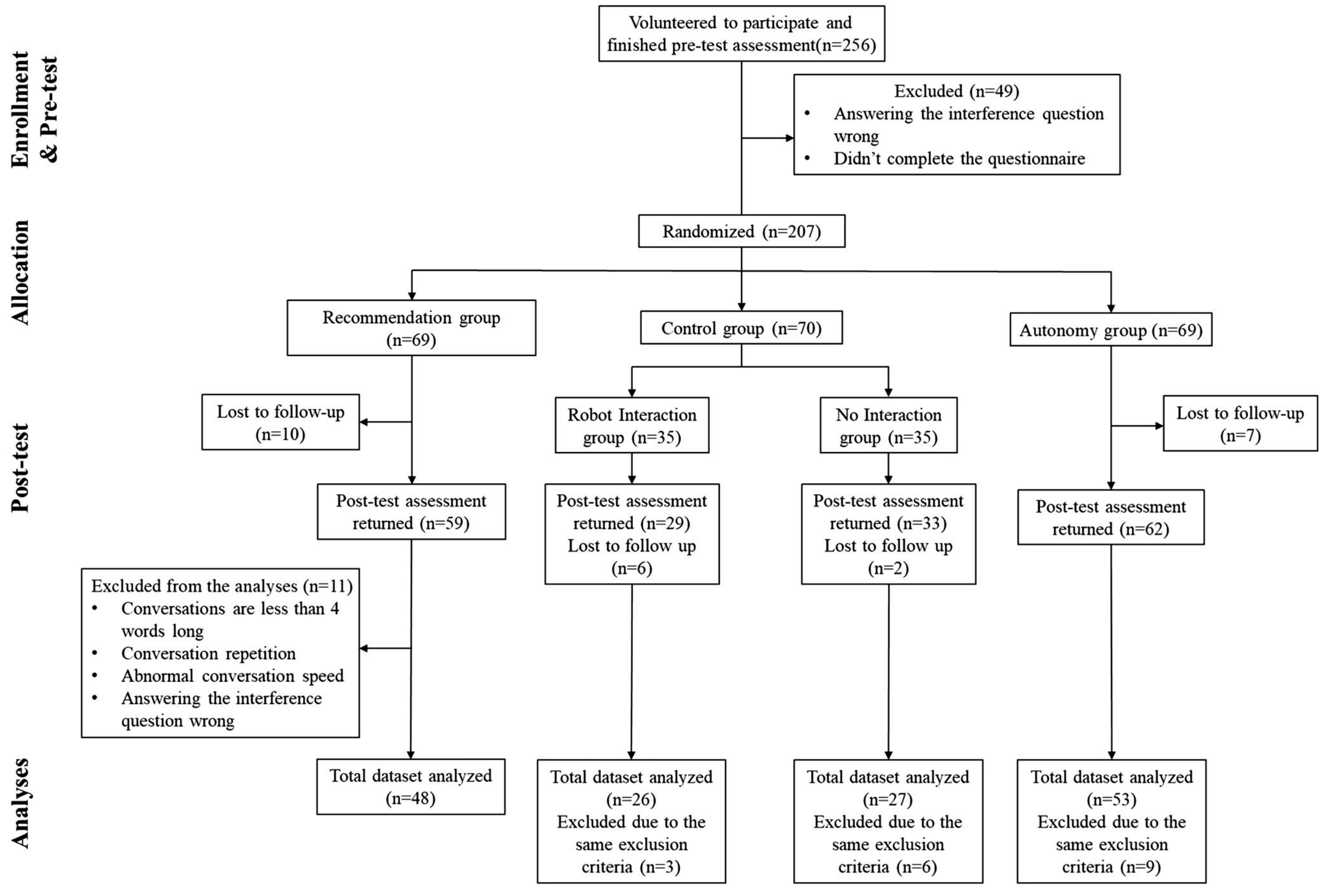

Figure 1. Procedure flowchart for sub-study 1.

states. We allocated the remaining participants to the gratitude exercise.

This study then allowed participants in the autonomy group to select their preferred PPIs and expected them to achieve better outcomes than the control group (Auyeung & Mo, 2019). In the control group, we randomly assigned PPIs to participants.

This study asked participants to choose a 10-minute time slot between 6 p.m. and 3 a.m. every day for six consecutive days to interact with a chatbot. During this period, they completed six sessions of positive mental exercises. At the end of the sixth day, participants were asked to complete a post-test questionnaire. Data from sub-study 1 and 2 were partially presented orally as a late-breaking-work short report at the 2022 Chinese CHI conference.

For statistical analysis, a paired-sample t-test was used to assess the intervention's efficacy within each group. We then conducted the mixed-design analysis of variance (ANOVA) to examine the interaction effects between the group and time, and applied the Bonferroni correction for multiple comparisons to reduce Type I errors. To assess the significance of the intervention's impact on mental health, we calculated effect sizes.

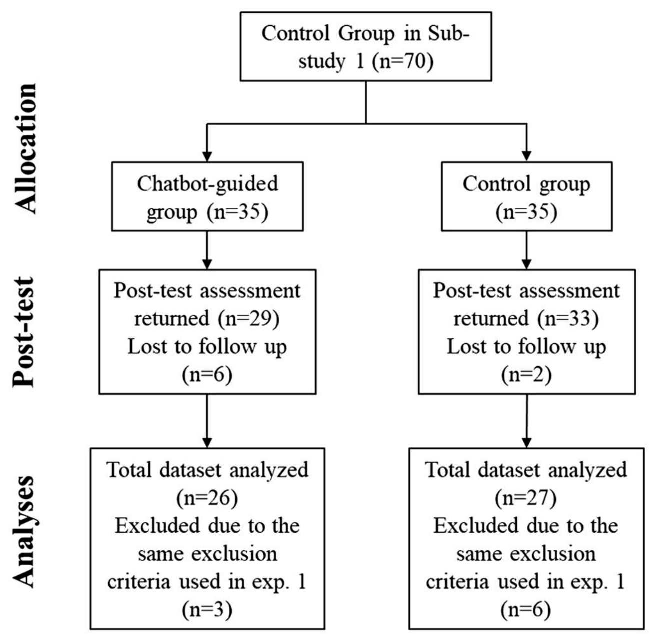

Figure 2. Procedure flowchart for sub-study 2.

2.2. Sub-study 2

After completing sub-study 1, we began sub-study 2. In this phase, 70 participants from the control group of sub-study 1 were randomly re-assigned to either the chatbot-guided or control groups (Figure 2). We did not add any new participants during this phase. By the study's conclusion, 53 participants had finished the entire experiment. In the chatbot-guided group, the chatbot directed participants through

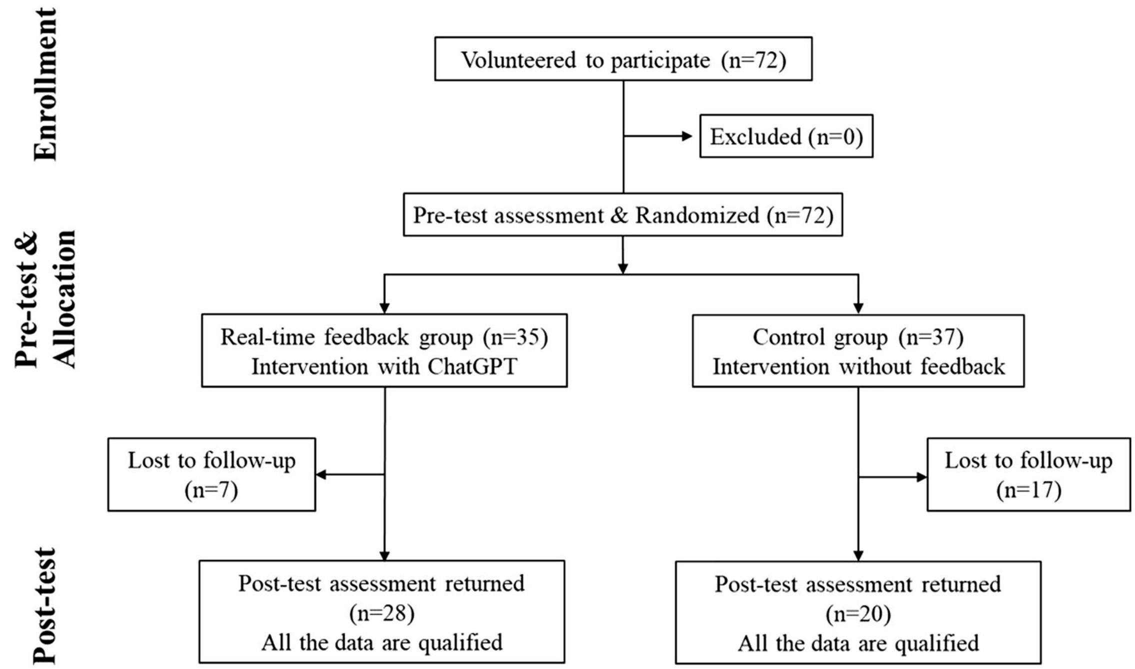

Figure 3. Procedure flowchart for sub-study 3.

the PPIs using an adaptive multi-round dialogue. In contrast, the chatbot in the control group provided an online PPI, giving participants detailed, one-off instructions to carry out the intervention on their own.

Given the limited sample size, it was essential to assess the normality of each variable using the Kolmogorov-Smirnov test. For those variables that were normally distributed, paired-sample t-test and independent sample t-test were applied to examine differences in score changes within and between groups, respectively. For variables that did not follow a normal distribution, we employed two non-parametric tests, the paired Wilcoxon signed-rank test to evaluate changes within groups and the Mann-Whitney U test to analyze differences in score changes between groups.

2.3. Sub-study 3

Sub-study 3 recruited a new group of participants, randomly allocated into two sections. Participants were reimbursed 30 RMB (approximately 5 USD). The real-time feedback group comprised 28 individuals (46% male, average age: 21.14), while the control group included 22 participants (36% male) (Figure 3).

In the two-week experiment, the chatbot in both groups offered a variety of PPIs, including TGT, BPS, Positive Introduction, Character Strengths Assessment, Know-How of Strengths, Tree of Positive Relationships, Positive Legacy, and Count Your Blessings, all adapted from the *Positive Psychotherapy: Clinician Manual* (Rashid & Seligman, 2018). For the real-time feedback group, after each exercise, the chatbot provided immediate and encouraging feedback based on the participant's input during the intervention (Figure 4). Conversely, participants in the control group received online PPI guidance through the chatbot but did not receive any post-intervention response. The data analysis procedure followed the same process as in sub-study 2.

2.4. Chatbot design

There are three types of chatbots: The rule-based chatbot responds based on a fixed, predefined set of rules (Adamopoulou & Moussiades, 2020); The retrieval-based chatbot selects an appropriate response to the current conversation from a repository, using response selection algorithms (Wu et al., 2016); The generative chatbot creates answers based on current and previous user messages, employing machine learning or deep learning techniques.

We leveraged the Baidu UNIT Dialogue Platform (Baidu Inc., China) to develop a retrieval-based chatbot for the first and second sub-studies. All PPI guidance dialogs were written in advance, and UNIT determined the most appropriate route using NLP technology after interacting with a user.

In the third sub-study, we sought to assess the importance of timely, relevant, and encouraging feedback. However, the user experience might suffer if the chatbot fails to offer appropriate feedback. Considering that a retrieval-based chatbot can only generate responses from a finite set of predefined options, and given the unpredictability of topics and content users might broach during positive psychology exercises, designing a retrieval-based chatbot capable of providing proper feedback posed significant challenges. As a solution, we turned to the generative chatbot. More specifically, we used the ChatGPT-3.5-Turbo API by OpenAI LP, US, to create a generative pre-trained transformer (GPT)-based chatbot for the third sub-study. (Note: At the time of the study, GPT-3.5-Turbo was the most recent version available. However, by the time of this article's submission, GPT 4.0 had been released.)

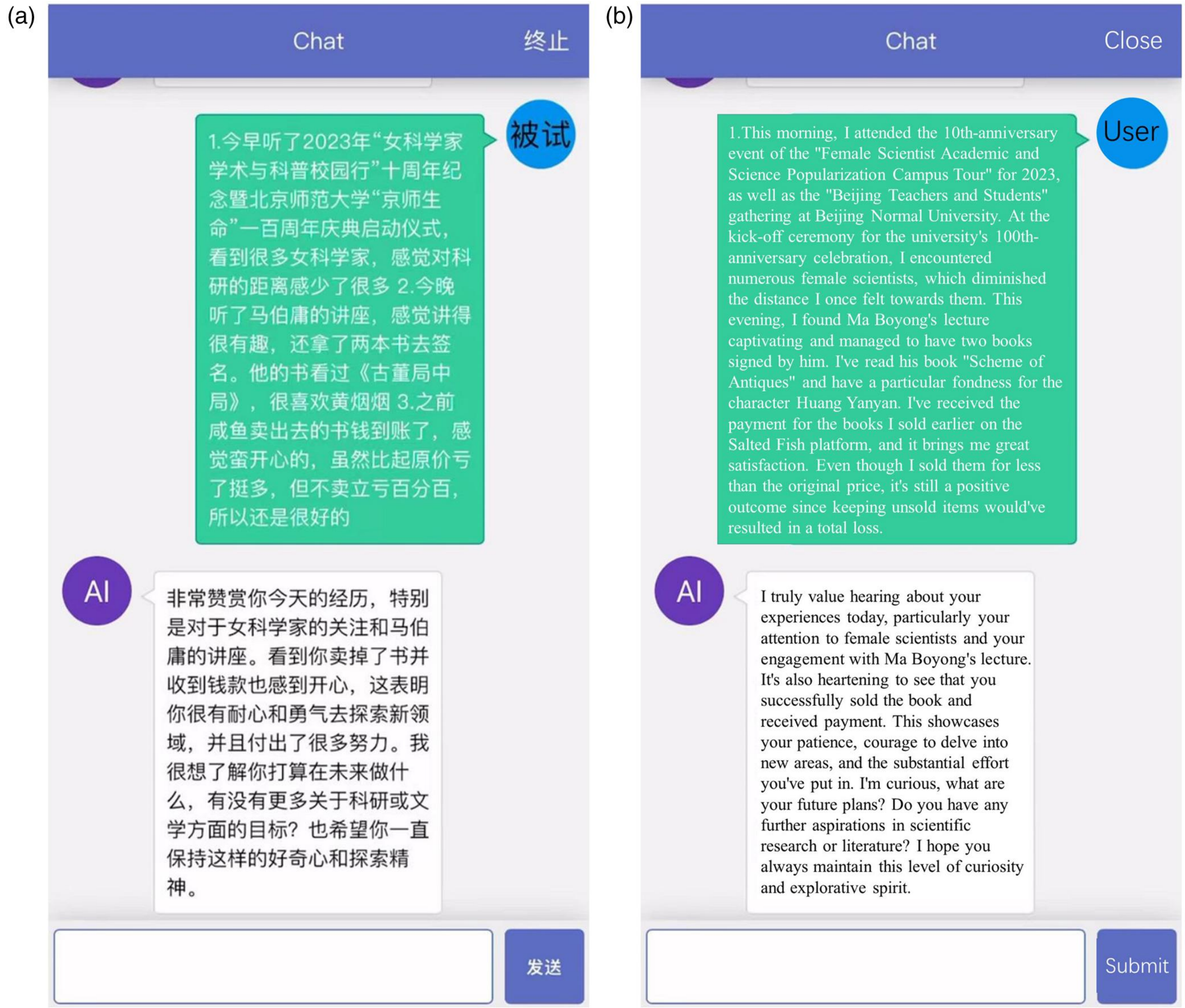

Figure 4. (a) An example of the chatbot user interface and the real-time feedback generated by GPT-based chatbot upon completion of the TGT exercise by the user. (b) The English translation of the example.

GPT-3.5-Turbo is a subclass of GPT-3, which is a large language model (LLM) with 175 billion parameters spread across 96 layers. GPT-3 has been trained on a vast corpus of 499 billion tokens from web content, utilizing a deep learning architecture known as the transformer (Dale, 2021). The transformer serves as a neural network component specifically tailored for processing sequential data like text and predicting subsequent tokens. Researchers have demonstrated that GPT-based (or LLM-based) chatbots have the ability to generate conversational sentences nearly indistinguishable from human-generated text. Additionally, studies indicate that by tweaking the input (or prompt) for the model, GPT-based chatbots can alter their tone or even convey more empathy (Lee et al., 2022).

The generative chatbot offers significant potential to guide users through psychological interventions. Given the importance of making users feel genuinely heard and understood in such contexts, the responsiveness of generative chatbots (Deng & Lin, 2023), coupled with their ability to detect emotional states and provide empathetic feedback, position them as viable alternatives to human therapists, especially when more costly human interactions are not feasible. Additionally, generative chatbots can address a vast spectrum of topics and, as we have observed, sometimes display knowledge beyond human capabilities. Also, steering GPT's output via prompt engineering is intuitive (Zamfirescu-Pereira et al., 2023), making it easy for individuals with limited technical know-how to design and implement such tools.

Twelve out of the fourteen PPIs in sub-study 3 involved expressive writing. We added prompts to guide the chatbot in providing encouraging feedback, aiming to make participants feel valued and eager to engage in the exercise the following day. Examples of prompts used in sub-study 3 include: “Express admiration or empathy, acknowledge the effort I've made, ask for more details, and express concern about my future plans,” “Recognize my plan and encourage me to persevere,” “Show empathy for my experiences and

Table 1. Data analysis results for sub-study 1.

		Intra-group comparisons												Between-group comparisons		
		Recommendation				Control				Autonomy						
		N	M	SD	t	N	M	SD	t	N	M	SD	t	F	η2	Post hoc
PHQ-9	Pre-test	48	17.58	6.2	3.74***	52	17.69	4.44	2.69**	53	16.53	5.59	1.52	1.25	0.02	
	Post-test	48	15.19	4.94		53	16.58	4.62		52	15.35	5.14				
GAD-7	Pre-test	48	16.77	7.31	3.04**	53	17.02	7.05	2.56*	53	14.98	5.92	0.8	0.85	0.01	–
	Post-test	48	14.35	6.69		53	15.15	5.91		53	14.32	6.65				
PANAS-P	Pre-test	48	32.81	6.74	−2.47*	53	33.53	5.85	−1.18	53	34.74	6.92	−0.83	0.98	0.01	–
	Post-test	48	34.88	7.1		53	34.57	6.06		52	35.67	5.91				
PANAS-N	Pre-test	48	23.46	8.08	5.16***	52	22.31	7.86	3.28**	53	20.91	8.01	1.11	5.68**	0.07	Recom.>Auto
	Post-test	48	19.44	6.73		52	19.83	6.41		52	19.63	7.88				
SWLS	Pre-test	48	19.06	5.61	−3.81***	53	21.04	5.12	−2.07*	53	22.21	7.25	−2.63*	0.85	0.01	–
	Post-test	48	21.92	5.1		53	22.11	4.44		53	23.83	6.85				
SVS	Pre-test	47	33.02	8.19	−1.86	53	33.19	6.36	−1.38	53	34.57	7.5	0.21	4.13*	0.05	Cont.>Auto
	Post-test	48	34.38	6.89		53	34.26	5.9		53	34.38	8.35				

p value: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

respond positively" and "This is a letter of forgiveness. Express your understanding and empathy, and aim to make me feel comforted."

2.5. Questionnaire

The study employed the following questionnaires: The Patient Health Questionnaire-9 (PHQ-9) is a self-reported depression screening tool aligning with DSM-IV guidelines. This nine-item questionnaire employs a four-point scoring system and showed high reliability in this study (Kroenke et al., 2001), reflected in Cronbach's alpha values of 0.873 and 0.851 for the two groups of participants (recruited before sub-study 1 and 3). The second tool, the Generalized Anxiety Disorder-7 (GAD-7), measures generalized anxiety disorder and symptom severity. Consisting of 7 items, it also uses a four-point scoring system with a total score range of 0–21(Spitzer et al., 2006). The GAD-7 exhibited strong reliability as well, with Cronbach's alpha values of 0.924 and 0.950 for the two groups of participants, respectively.

The Positive Affect and Negative Affect Scale (PANAS) is a 20-item self-report tool incorporating a five-point scoring system (Watson et al., 1988). It evaluates current emotional states via ten positive and ten negative descriptors. The scale exhibited robust internal consistency in the two groups with Cronbach's alpha values for positive affect at 0.821 and 0.825, and for negative affect at 0.898 and 0.919. The Satisfaction With Life Scale (SWLS) is a 5-item tool assessing life satisfaction (Diener et al., 1985). It gauges agreement with item descriptions using a 7-point scale, with total scores ranging from 5 to 35. In the two groups, the scale achieved Cronbach's alpha values of 0.874 and 0.867, respectively, indicating strong internal consistency.

The Scales of Psychological Well-being (PWB) is an 18-item self-report tool that assesses well-being across six facets: self-acceptance, positive relations, autonomy, environmental mastery, purpose in life, and personal growth (Ryff & Keyes, 1995). Only participants in sub-study 3 completed this scale, yielding a satisfactory Cronbach's alpha value of 0.797, indicating consistent internal reliability. The Subjective Vitality Scale (SVS) evaluates perceived vitality and serves as a well-being indicator of personal well-being (Ryan & Frederick, 1997). This scale was completed solely by participants in sub-study 1 and 2, achieving a Cronbach's alpha value of 0.865.

3. Results

3.1. Sub-study 1

No significant differences were found in the pre-test scores between groups except for the SWLS score, which was markedly lower in the recommendation group as compared to the self-selection group ($p = 0.031$). The post-intervention results revealed the most substantial improvements in the recommendation group, indicating a significant reduction in depression, anxiety, and negative affect scores and a considerable increase in positive affect and life satisfaction scores (see Table 1). For the autonomy group, there was a decline in negative affect and an upswing in positive affect; however, only the increase in life satisfaction scores achieved statistical significance. The control group also showed significant improvements: depression, anxiety, and negative affect scores decreased markedly, whereas SWLS scores increased noticeably. In summary, regardless of the method for PPI selection, Chat-PPI has proven to be effective in reducing negative emotions and enhancing emotional well-being.

Further Bonferroni correction and post hoc analysis revealed that the recommendation group saw a greater decrease in negative affect than the autonomy group, and the control group's change in subjective vitality significantly differed from that in the autonomy group. The differences in other paired comparisons were not significant.

3.2. Sub-study 2

The normality test showed deviations from a normal distribution in the changes of SWLS scores ($p < 0.001$) in the chatbot-guided interaction group and the post-test scores of PANAS-P ($p = 0.032$), depression ($p = 0.002$), and anxiety ($p = 0.022$) in the control group. Examination of pre-test scores revealed a high degree of participant homogeneity across groups that only the SWLS scores were significantly lower in the chatbot-guided group than in the control group ($p = 0.009$). Within-group analysis demonstrated that the

Table 2. Data analysis results for sub-study 2.

Intra-group comparisons											
		Chatbot-guided				Control				Between-group comparisons	
		N	*M*	*SD*	*t*	*N*	*M*	*SD*	*t or Z*	*t or Z*	*cohen's d*
PHQ-9	Pre-test	26	18.27	4.57	2.53*	26	17.12	4.32	1.17	−1.73†	–
	Post-test	26	16.50	4.89		27	16.67	4.44			
GAD-7	Pre-test	26	17.77	7.05	2.40*	27	16.30	7.10	1.36	−1.23†	–
	Post-test	26	15.58	5.91		27	14.74	6.00			
PANAS-P	Pre-test	26	33.96	4.85	−0.30	27	33.11	6.75	−1.54	−0.31	0.09
	Post-test	26	34.35	7.27		27	34.78†	4.74			
PANAS-N	Pre-test	26	23.12	7.44	2.99**	26	21.50	8.32	1.55	−1.09	0.30
	Post-test	25	19.12	5.91		27	20.48	6.88			
SWLS	Pre-test	26	19.19	4.20	−3.96***	27	22.81	5.37	0.46	−2.39†*	–
	Post-test	26	21.73	3.76		27	22.48	5.06			
SVS	Pre-test	26	31.85	7.02	−0.81	27	34.48	5.47	−1.14	0.45	0.12
	Post-test	26	32.81	5.89		27	35.67	5.66			

p value: $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.
†non-parametric test.

Table 3. Data analysis results for sub-study 3.

		Intra-group comparisons									
		Real-time feedback (ChatGPT)				Control				Between-group comparisons	
		N	*M*	*SD*	*t or Z*	*N*	*M*	*SD*	*t or Z*	*t or Z*	*cohen's d*
PHQ-9	Pre-test	28	8.57	5.04	2.15*	20	8.75	5.22	−1.00†	0.97	0.28
	Post-test	28	7.07	3.50		19	7.53†	3.29			
GAD-7	Pre-test	28	8.43	6.96	−0.80†	20	9.75	7.03	0.38	0.16	0.05
	Post-test	28	7.68†	5.95		20	9.25	4.72			
PANAS-P	Pre-test	28	28.79	6.44	−1.17	20	31.6	6.55	0.88	−1.04	0.30
	Post-test	27	30.07	5.38		18	30.89	4.39			
PANAS-N	Pre-test	28	21.11	8.18	−0.35†	20	24.35	8.68	−0.26	0.04	0.01
	Post-test	28	20.18†	7.35		20	24.9	8.35			
SWLS	Pre-test	28	16.61	5.63	−2.46*	20	19.65	4.49	0.55	−2.02*	0.59
	Post-test	28	18.96	5.93		20	19.05	6.83			
PWB	Pre-test	28	85.46	13.58	−3.81***	20	87.6	10.09	0.32	−2.52†*	–
	Post-test	27	93.48	8.02		18	88.17	10.26			

Pre-post t-test p value. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$, †non-parametric test.

chatbot-guided group exhibited significant decreases in depression, anxiety, and negative affect scores, along with a substantial increase in life satisfaction and subjective vitality scores (see Table 2). However, despite the trends in the control group aligning with expectations, they did not reach statistical significance, suggesting a less pronounced effect from the intervention.

When comparing pre-test to post-test changes, our data revealed that life satisfaction in the chatbot-guided group improved significantly more than in the control group ($p = 0.017$). Although the rest of the outcomes were as expected, their differences did not reach statistical significance.

3.3. Sub-study 3

The scores for GAD-7 ($p = 0.007$), PANAS-N ($p = 0.033$), and the change in PWB ($p = 0.018$) in the real-time feedback group, along with the PHQ-9 post-test score ($p = 0.001$) in the control group, deviated from a normal distribution. Moreover, there was no significant difference in pre-test scores on each scale between different groups. Within-group analysis demonstrated a significant improvement in mental state for the real-time feedback group, evidenced by considerable decreases in depression scores and significant increases in SWLS and PWB scores post-intervention (see Table 3). Conversely, the control group exhibited no significant changes, suggesting a less pronounced intervention efficacy. A comparison between groups showed a significantly greater improvement effect in the real-time feedback group on SWLS and PWB scores compared to the control group. Therefore, our overall data support our initial hypothesis.

4. Discussions

4.1. Principle findings

The findings of this study align with our expectations and consistently reinforce, through three sub-studies, the substantial promise of Chat-PPI. These results further strengthen the potential of chatbots within the broader domain of mental health research. The randomized control test confirmed the three proposed hypotheses regarding the mechanisms determining the effectiveness of PPIs. In particular, providing personalized recommendations (H1), engaging users in adaptive multi-round dialogues (H2), and offering real-time empathetic and encouraging feedback based on user-provided content during practice (H3) all emerged as crucial mechanisms through which chatbots can effectively facilitate PPIs. However, we did not find that PPIs in the self-selection condition outperformed the PPIs in the random-selection condition, as previous studies found (Heintzelman et al., 2023), and the difference between personal recommendation and random selection did not reach statistical significance.

Our results echo previous studies suggesting that anthropomorphic and interactive chatbots can alleviate the monotony associated with long-term repeated interventions, thus fostering sustained user engagement (Adam et al., 2021). Moreover, these chatbots can exhibit empathy, fostering a sense of understanding in users during the psychological intervention process, thereby enhancing the intervention's efficacy. These findings might also indirectly explain the insignificant results of some previous Chat-PPI studies that did not emphasize these three characteristics.

Before the study's completion, we conducted unstructured interviews to delve deeper into the participants' experiences with the chatbots. Many participants expressed that the chatbot's responses evoked feelings of surprise and novelty. They emphasized the importance of feeling understood as pivotal in enhancing their acceptance and engagement with the chatbot-based intervention. Since GPT-based chatbots can generate text in a human-like manner, mimicking the interaction between users and psychotherapists (Uludag, 2023), five participants indicated that the anthropomorphic responses felt so authentic that they found it difficult to distinguish whether they were AI-generated. These qualitative insights reinforce our quantitative findings and underscore the critical importance of interaction in the design of chatbots. Notably, the warm and empathetic chatbot responses were designed by employing simple prompts. Therefore, our results underscore the GPT-based chatbot's convenience and user-friendliness, especially for designers with limited technical expertise. However, future research should investigate advanced prompt engineering and model fine-tuning methods to ensure an even more enhanced and consistent user experience.

4.2. Limitations and suggestions

4.2.1. Experiment design and control

While most of our findings corroborated our hypotheses, suggesting that Chat-PPI could effectively boost well-being and mitigate negative emotions, certain discrepancies between the experimental and control groups did not reach statistical significance. We suspect this lack of sufficient statistical power is due to certain aspects of our experimental design, aspects that could potentially be modified or improved in future research:

Firstly, managing external variables in chatbot-based studies can be complex due to the unpredictable nature of user interactions with both generative and retrieval chatbots. The unique content in each conversation, coupled with the multitude of possible paths in multi-round dialogues, amplifies experimental intricacy. This complexity challenges the foundational assumptions of causal inference and complicates the replication of experiments. Psychological studies have been enduring a persistent issue of low statistical power (Wagenmakers et al., 2015). Despite our sample size of 328 participants, the group sizes did not adequately compensate for this inherent randomness. Future research should involve a larger participant pool or collecting more variables from each participant, as suggested by previous studies (Dumas-Mallet et al.,2017).

Secondly, the novelty of technologies, defined as a stimulus that is unknown or unfamiliar to an individual (Skavronskaya et al., 2020), can potentially compromise experimental results. Research on new technologies has underscored that the effect of novelty could lead to experimental bias (Mirnig et al., 2020). Throughout our experimental process, we observed that initial interactions with the chatbot elicited amplified positive emotions among participants. The various degrees of participants' familiarity with and attitudes toward technology can also influence their engagement with Chat-PPIs, subsequently increasing the randomness that diminishes statistical power (Baños et al., 2014). Allowing participants more time to familiarize themselves with the chatbot before the study might help reduce disparities from technological familiarity. Future studies should consider including technology familiarity as a control variable to eliminate its potential influence.

Thirdly, ensuring participant compliance in online psychological experiments is a challenge. Several participants diverged from the predefined experimental procedures or schedules in our study. High attrition rates, especially among male participants, further compromised the reliability of our data. Given these challenges, future studies should explore strategies to enhance participant motivation and adherence. One potential solution could be the introduction of a monetary incentive scheme.

Fourthly, to manage participant behavior, reduce experimental costs, and minimize attrition, the duration of the experiments in this study was kept to one or two weeks. Although the optimal length for PPIs is still a subject of debate (Koydemir et al., 2021), a significant amount of research suggests that extended PPI programs yield greater improvements in well-being and strengths (Carr et al., 2021). Consequently, the short duration of our experiment limited the effectiveness of our intervention. Future studies should consider prolonging the intervention period to obtain more substantial outcomes.

Finally, it is worth noting that our choice of an active control group (that is, a non-trivial control group) in a randomized controlled trial, as suggested by earlier studies (Van Zyl et al., 2019), might have led to decreased statistical power relative to the use of a no-treatment control group (that is, a group receiving no intervention or placed on a waitlist). In our study, all control groups were assigned to technology-assisted PPIs, which had already been empirically supported in prior research. Thus, while all our data support the effectiveness of Chat-PPI, the degree of improvement provided by our proposed methods might not achieve statistical significance.

4.2.2. The choice of PPIs

Programs that incorporate multiple PPIs usually display greater efficacy, especially in reducing depressive symptoms, compared to those centered around a single PPI (Carr et al., 2021). However, such programs often administer a range of PPIs to participants without adequately differentiating their distinct effects, contexts, and durations (Schueller & Parks, 2012). It is important to consider that PPIs exhibit considerable differences in their effectiveness and durations (Khanna & Singh, 2019). For instance, interventions promoting savoring, optimism, and hope have demonstrated more significant impacts than other types of PPIs (Carr et al., 2021). Furthermore, participants in a low mood may struggle to identify three positive daily events for the TGT exercise, and there have been reports that excessive repetition of TGT throughout the week can lead to user distress (Lyubomirsky, 2008).

Despite these distinctions among PPIs, current research offers scant guidance on how to implement a variety of PPIs concurrently. As a result, this study, mirroring previous research, did not distinguish between the different PPIs except for the personalized recommendation implemented in sub-study 1. Given the crucial role of integrating multiple PPIs in enhancing participant engagement and boosting the efficacy of the intervention, future research must explore this area in greater depth.

4.2.3. Chatbot-design

Chatbots serve as a pioneering platform for psychological interventions, where their effectiveness and usability are greatly dependent on the design of their dialogues and interactions. In sub-studies 1 and 2, we chose a retrieval-based chatbot to effectively manage the psychological intervention process. However, the constrained interactiveness resulted in users losing interest after just a few days of usage. This observation is consistent with previous studies, which indicate that the limited responsiveness of chatbots can trigger user frustration, miss therapeutic opportunities, and even provoke inappropriate responses (Miner et al., 2016).

Conversely, our third sub-study leveraged GPT to offer empathetic feedback and highlighted that generative chatbots bring superior advantages to psychological interventions compared to retrieval-based chatbots in several respects. With thoughtful design, generative chatbots can offer personalized recommendations, adaptable dialogues, and immediate feedback, thereby amplifying user engagement and fostering empathy. Surprisingly, contrary to our initial presumptions, the development of retrieval-based chatbots often demands more developer time and resources compared to the creation of GPT-based chatbots via prompt engineering techniques. Furthermore, implementing a GPT-based chatbot sometimes requires less expertise in programming and machine learning. User-friendly Python frameworks, such as the Langchain toolkit (Langchain, Delaware, USA), present a more enticing option for psychological researchers who do not have an extensive technical background.

Another interesting finding from our study was that GPT-based chatbots could produce responses more swiftly and exhibited an understanding of a broader range of topics than their human counterparts. In a preliminary investigation before sub-study 3, we engaged human therapists to offer real-time responses. Nonetheless, they often required significantly more time to construct their responses, occasionally leading to participant impatience. Moreover, human therapists sometimes demonstrated less understanding of the topics brought up by participants when compared to the GPT-based chatbots. However, a direct comparison of Chat-PPI and human-based interventions was beyond the scope of this study and necessitated further investigation in future research.

Despite these advantages, concerns have been raised about potential harmful outcomes arising from unregulated and unchecked dialogue content (Abd-Alrazaq et al., 2020). The complexity and variability of human language sometimes make it difficult to understand (Chow et al., 2023). Moreover, a false response by GPT can be particularly dangerous in medical scenarios because the errors can be subtle and convincing (Lee et al., 2023). However, the existing literature has not adequately supported the safety of chatbots (Van Dis et al., 2023), limiting their usage in psychological interventions. Nevertheless, consistent with several studies that have confirmed the accuracy of GPT-generated outputs (Johnson et al., 2023), our exhaustive analysis of all chatbot-user dialogues recorded in this study did not reveal any inappropriate responses. This result, however, is not definitive, and further research is required to ascertain the safety of chatbots in psychological research.

Moreover, as noted earlier, due to the complexities involved in using GPT for multi-round dialogue design, we employed a retrieval-based chatbot for sub-studies 1 and 2 and used a GPT-based chatbot only to deliver one-step real-time feedback for sub-study 3. Following sub-study 3, our team continued to develop a GPT-based chatbot that could lead users through a multi-round psychological intervention. However, controlling chatbots for such a comprehensive psychological intervention process remains challenging, even when utilizing advanced fine-tuning techniques (Ouyang et al., 2022). For future research, we propose the development of hybrid chatbots, wherein a rule-based or retrieval-based model is used to direct users through the therapeutic and intervention processes, while a generative chatbot (including GPT-based models) facilitates interaction and delivers real-time feedback. Our preliminary trials using this approach have shown promising results.

5. Conclusion

Chatbots hold significant potential in PPI research, yet this field has been largely unexplored in academic literature, with very few rigorous randomized control trial experiments conducted. Apart from confirming the effectiveness of Chat-PPI and endorsing the use of technology-assisted PPIs—which have been found to be less effective than traditional human-to-human PPIs (Koydemir et al., 2021)—this study pioneers the exploration of the mechanisms that can enhance the effectiveness of chatbots in PPI. Our findings further reinforce that interactive elements, such as personalized recommendations, multi-round dialogues, and real-time feedback, are key influential factors for chatbot-assisted PPIs. In terms of chatbot design, this research not only utilizes retrieval-based chatbots but also innovatively incorporates GPT-based chatbots to guide users through psychological intervention. This constitutes a groundbreaking step, not only in PPI but also in the wider field of psychological intervention studies (Pandey & Sharma, 2023). Thus, our experiences with generative chatbots—including the design challenges we faced and our investigation into dialogue safety—offer valuable insights for future researchers.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Beijing Normal University at Zhuhai Researcher Activation Fund (Grant No. 310432101), Shenzhen Key Laboratory of Next Generation Interactive Media Innovative Technology (Grant No. ZDSYS20210623092001004), Chinese National Social Science Foundation (Grant No.BBA210042), the Shenzhen Key Research Base of Humanities, Social Sciences for People's Well-being Benchmarking Study (Grant No. 202003) and the Guangdong Digital Mental Health and Intelligent Generation Laboratory (Grant No. 2023WSYS010).

ORCID

Ivan Liu http://orcid.org/0000-0001-8171-9356

References

Abd-Alrazaq, A. A., Rababeh, A., Alajlani, M., Bewick, B. M., & Househ, M. (2020). Effectiveness and safety of using chatbots to improve mental health: Systematic review and meta-analysis. *Journal of Medical Internet Research*, *22*(7), e16021. https://doi.org/10.2196/16021

Adam, M., Wessel, M., & Benlian, A. (2021). AI-based chatbots in customer service and their effects on user compliance. *Electronic Markets*, *31*(2), 427–445. https://doi.org/10.1007/s12525-020-00414-7

Adamopoulou, E., & Moussiades, L. (2020). An overview of chatbot technology. IFIP International Conference on Artificial Intelligence Applications and Innovations (pp. 373–383).

Ahmad, R., Siemon, D., Gnewuch, U., & Robra-Bissantz, S. (2022). Designing personality-adaptive conversational agents for mental health care. *Information Systems Frontiers*, *24*(3), 923–943. https://doi.org/10.1007/s10796-022-10254-9

Auyeung, L., & Mo, P. K. H. (2019). The efficacy and mechanism of online positive psychological intervention (PPI) on improving well-being among Chinese university students: A pilot study of the best possible self (BPS) intervention. *Journal of Happiness Studies*, *20*(8), 2525–2550. https://doi.org/10.1007/s10902-018-0054-4

Ayers, J. W., Poliak, A., Dredze, M., Leas, E. C., Zhu, Z., Kelley, J. B., Faix, D. J., Goodman, A. M., Longhurst, C. A., Hogarth, M., & Smith, D. M. (2023). Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA Internal Medicine*, *183*(6), 589–596. https://doi.org/10.1001/jamainternmed.2023.1838

Baños, R. M., Etchemendy, E., Farfallini, L., García-Palacios, A., Quero, S., & Botella, C. (2014). EARTH of well-being system: A pilot study of an information and communication technology-based positive psychology intervention. *The Journal of Positive Psychology*, *9*(6), 482–488. https://doi.org/10.1080/17439760.2014.927906

Boehm, J. K., & Kubzansky, L. D. (2012). The Heart's content: The association between positive psychological well-being and cardiovascular health. *Psychological Bulletin*, *138*(4), 655–691. https://doi.org/10.1037/a0027448

Boucher, E. M., Harake, N. R., Ward, H. E., Stoeckl, S. E., Vargas, J., Minkel, J., Parks, A. C., & Zilca, R. (2021). Artificially intelligent chatbots in digital mental health interventions: A review. *Expert Review of Medical Devices*, *18*(suppl 1), 37–49. https://doi.org/10.1080/17434440.2021.2013200

Carr, A., Cullen, K., Keeney, C., Canning, C., Mooney, O., Chinseallaigh, E., & O'Dowd, A. (2021). Effectiveness of positive psychology interventions: A systematic review and meta-analysis. *The Journal of Positive Psychology*, *16*(6), 749–769. https://doi.org/10.1080/17439760.2020.1818807

Chaves, A. P., & Gerosa, M. A. (2021). How should my chatbot interact? A survey on social characteristics in human-chatbot interaction design. *International Journal of Human–Computer Interaction*, *37*(8), 729–758. https://doi.org/10.1080/10447318.2020.1841438

Cheng, S., Chang, C., Chang, W., Wang, H., Liang, C., Kishimoto, T., Chang, J. P., Kuo, J. S., & Su, K. (2023). The now and future of ChatGPT and GPT in psychiatry. *Psychiatry and Clinical Neurosciences*, *77*(11), 592–596. https://doi.org/10.1111/pcn.13588

Chow, J. C., Sanders, L., & Li, K. (2023). Design of an educational chatbot using artificial intelligence in radiotherapy. *AI*, *4*(1), 319–332. https://doi.org/10.3390/ai4010015

Dale, R. (2021). GPT-3: What's it good for? *Natural Language Engineering*, *27*(1), 113–118. https://doi.org/10.1017/S1351324920000601

Deng, J., & Lin, Y. (2023). The benefits and challenges of ChatGPT: An overview. *Frontiers in Computing and Intelligent Systems*, *2*(2), 81–83. https://doi.org/10.54097/fcis.v2i2.4465

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment*, *49*(1), 71–75. https://doi.org/10.1207/s15327752jpa4901_13

Dumas-Mallet, E., Button, K. S., Boraud, T., Gonon, F., & Munafò, M. R. (2017). Low statistical power in biomedical science: A review of three human research domains. *Royal Society Open Science*, *4*(2), 160254. https://doi.org/10.1098/rsos.160254

Emmons, R. A., & McCullough, M. E. (2003). Counting blessings versus burdens: An experimental investigation of gratitude and subjective well-being in daily life. *Journal of Personality and Social Psychology*, *84*(2), 377–389. https://doi.org/10.1037//0022-3514.84.2.377

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), e19. https://doi.org/10.2196/mental.7785

Fredrickson, B. L. (2001). The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions. *The American Psychologist*, *56*(3), 218–226. https://doi.org/10.1037//0003-066x.56.3.218

Greer, S., Ramo, D., Chang, Y.-J., Fu, M., Moskowitz, J., & Haritatos, J. (2019). 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 and uHealth*, *7*(10), e15018. https://doi.org/10.2196/15018

Grové, C. (2021). Co-developing a mental health and wellbeing chatbot with and for young people. *Frontiers in Psychiatry*, *11*. https://doi.org/10.3389/fpsyt.2020.606041

Haristiani, N. (2019). Artificial Intelligence (AI) chatbot as language learning medium: An inquiry. *Journal of Physics: Conference Series*, *1387*(1), 012020. https://doi.org/10.1088/1742-6596/1387/1/012020

Heintzelman, S. J., Kushlev, K., & Diener, E. (2023). Personalizing a positive psychology intervention improves well-being. *Applied Psychology: Health and Well-Being*, *15*(4), 1271–1292. https://doi.org/10.1111/aphw.12436

Howells, A., Ivtzan, I., & Eiroa-Orosa, F. J. (2016). Putting the 'app' in happiness: A randomised controlled trial of a smartphone-based mindfulness intervention to enhance wellbeing. *Journal of Happiness Studies*, *17*(1), 163–185. https://doi.org/10.1007/s10902-014-9589-1

Jain, V., Rai, H., Subash, P., & Mogaji, E. (2023). The prospects and challenges of ChatGPT on marketing research and practices. *Emmanuel, the Prospects and Challenges of ChatGPT on Marketing Research and Practices*. https://doi.org/10.2139/ssrn.4398033

Jeong, S., Aymerich-Franch, L., Alghowinem, S., Picard, R. W., Breazeal, C. L., Park, H. W. (2023). A robotic companion for psychological well-being: A long-term investigation of companionship and therapeutic alliance. *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 485–494).

Johnson, D., Goodman, R., Patrinely, J., Stone, C., Zimmerman, E., Donald, R., Chang, S., Berkowitz, S., Finn, A., & Jahangir, E. (2023). Assessing the accuracy and reliability of AI-generated medical responses: An evaluation of the Chat-GPT model. *Research Square*. https://doi.org/10.21203/rs.3.rs-2566942/v1

Khanna, P., & Singh, K. (2019). Do all positive psychology exercises work for everyone? Replication of Seligman et al.'s (2005) interventions among adolescents. *Psychological Studies*, *64*(1), 1–10. https://doi.org/10.1007/s12646-019-00477-3

Koydemir, S., Sökmez, A. B., & Schütz, A. (2021). A meta-analysis of the effectiveness of randomized controlled positive psychological interventions on subjective and psychological well-being. *Applied Research in Quality of Life*, *16*(3), 1145–1185. https://doi.org/10.1007/s11482-019-09788-z

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, *16*(9), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Layous, K., Chancellor, J., Lyubomirsky, S., Wang, L., & Doraiswamy, P. M. (2011). Delivering happiness: Translating positive psychology intervention research for treating major and minor depressive disorders. *Journal of Alternative and Complementary Medicine*, *17*(8), 675–683. https://doi.org/10.1089/acm.2011.0139

Layous, K., Nelson, S. K., & Lyubomirsky, S. (2013). What is the optimal way to deliver a positive activity intervention? The case of writing about one's best possible selves. *Journal of Happiness Studies*, *14*(2), 635–654. https://doi.org/10.1007/s10902-012-9346-2

Lee, P., Bubeck, S., & Petro, J. (2023). Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. *The New England Journal of Medicine*, *388*(13), 1233–1239. https://doi.org/10.1056/NEJMsr2214184

Lee, Y.-J., Lim, C.-G., Choi, H.-J. (2022). Does GPT-3 generate empathetic dialogues? A novel in-context example selection method and automatic evaluation metric for empathetic dialogue generation. *Proceedings of the 29th International Conference on Computational Linguistics* (pp. 669–683).

Lim, S. M., Shiau, C. W. C., Cheng, L. J., & Lau, Y. (2022). Chatbot-delivered psychotherapy for adults with depressive and anxiety symptoms: A systematic review and meta-regression. *Behavior Therapy*, *53*(2), 334–347. https://doi.org/10.1016/j.beth.2021.09.007

Liu, H., Peng, H., Song, X., Xu, C., & Zhang, M. (2022). Using AI chatbots to provide self-help depression interventions for university students: A randomized trial of effectiveness. *Internet Interventions*, *27*, 100495. https://doi.org/10.1016/j.invent.2022.100495

Lomas, T., Waters, L., Williams, P., Oades, L. G., & Kern, M. L. (2021). Third wave positive psychology: Broadening towards complexity. *The Journal of Positive Psychology*, *16*(5), 660–674. https://doi.org/10.1080/17439760.2020.1805501

Loveys, K., Sagar, M., Pickering, I., & Broadbent, E. (2021). A digital human for delivering a remote loneliness and stress intervention to at-risk younger and older adults during the COVID-19 pandemic: Randomized pilot trial. *JMIR Mental Health*, *8*(11), e31586. https://doi.org/10.2196/31586

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. https://doi.org/10.1016/j.invent.2017.10.002

Lyubomirsky, S. (2008). *The how of happiness: A scientific approach to getting the life you want*. Penguin.

Masicampo, E. J., & Lalande, D. R. (2012). A peculiar prevalence of p values just below. 05. *Quarterly Journal of Experimental Psychology*, *65*(11), 2271–2279. https://doi.org/10.1080/17470218.2012.711335

Medeiros, L., Gerritsen, C., & Bosse, T. (2019). Towards Humanlike Chatbots Helping Users Cope with Stressful Situations. In N. T. Nguyen, R. Chbeir, E. Exposito, P. Aniorte, & B. Trawinski (Eds.), *Computational Collective Intelligence, Pt I*. (Vol. 11683, pp. 232–243). Springer International Publishing Ag. https://doi.org/10.1007/978-3-030-28377-3_19

Miner, A. S., Milstein, A., Schueller, S., Hegde, R., Mangurian, C., & Linos, E. (2016). Smartphone-based conversational agents and responses to questions about mental health, interpersonal violence, and physical health. *JAMA Internal Medicine*, *176*(5), 619–625. https://doi.org/10.1001/jamainternmed.2016.0400

Mirnig, A. G., Gärtner, M., Meschtscherjakov, A., Tscheligi, M. (2020). Blinded by novelty: A reflection on participant curiosity and novelty in automated vehicle studies based on experiences from the field. *Proceedings of Mensch Und Computer 2020* (pp. 373–381).

Nicolescu, L., & Tudorache, M. T. (2022). Human-computer interaction in customer service: The experience with AI chatbots—A systematic literature review. *Electronics*, *11*(10), 1579. https://doi.org/10.3390/electronics11101579

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., & Ray, A. (2022). Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, *35*, 27730–27744. https://proceedings.neurips.cc/paper_files/paper/2022/file/b1efde53be364a73914f58805a001731-Paper-Conference.pdf

Pandey, S., & Sharma, S. (2023). A comparative study of retrieval-based and generative-based chatbots using deep learning and machine learning. *Healthcare Analytics*, *3*, 100198. https://doi.org/10.1016/j.health.2023.100198

Park, C. L., Lim, H., Newlon, M., Suresh, D. P., & Bliss, D. E. (2014). Dimensions of religiousness and spirituality as predictors of well-being in advanced chronic heart failure patients. *Journal of Religion and Health*, *53*(2), 579–590. https://doi.org/10.1007/s10943-013-9714-1

Parks, A. C., & Biswas-Diener, R. (2013). Positive interventions: Past, present and future. In *Mindfulness, acceptance, and positive psychology: the seven foundations of well-being* (pp. 140–165). Context Press.

Paul, J., Ueno, A., & Dennis, C. (2023). ChatGPT and consumers: Benefits, pitfalls and future research agenda. In *International Journal of Consumer Studies* (Vol. 47, pp.1213–1225). Wiley Online Library. https://doi.org/10.1111/ijcs.12928

Rashid, T., & Seligman, M. P. (2018). *Positive psychotherapy: Clinician manual*. Oxford University Press.

Richards, D., & Richardson, T. (2012). Computer-based psychological treatments for depression: A systematic review and meta-analysis. *Clinical Psychology Review*, *32*(4), 329–342. https://doi.org/10.1016/j.cpr.2012.02.004

Ryan, R. M., & Frederick, C. (1997). On energy, personality, and health: Subjective vitality as a dynamic reflection of well-being. *Journal of Personality*, *65*(3), 529–565. https://doi.org/10.1111/j.1467-6494.1997.tb00326.x

Ryff, C., & Keyes, C. (1995). The structure of psychological well-being revisited. *Journal of Personality and Social Psychology*, *69*(4), 719–727. https://doi.org/10.1037/0022-3514.69.4.719

Rzepka, C., Berger, B., & Hess, T. (2022). Voice assistant vs. Chatbot–examining the fit between conversational agents' interaction modalities and information search tasks. *Information Systems Frontiers*, *24*(3), 839–856.

Schueller, S. M., Muñoz, R. F., & Mohr, D. C. (2013). Realizing the potential of behavioral intervention technologies. *Current Directions in Psychological Science*, *22*(6), 478–483. https://doi.org/10.1177/0963721413495872

Schueller, S. M., & Parks, A. C. (2012). Disseminating self-help: Positive psychology exercises in an online trial. *Journal of Medical Internet Research*, *14*(3), e1850. https://doi.org/10.2196/jmir.1850

Seligman, M. E., Steen, T. A., Park, N., & Peterson, C. (2005). Positive psychology progress: Empirical validation of interventions. *The American Psychologist*, *60*(5), 410–421. https://doi.org/10.1037/0003-066X.60.5.410

Siddique, S., & Chow, J. C. (2021). Machine learning in healthcare communication. *Encyclopedia*, *1*(1), 220–239. https://doi.org/10.3390/encyclopedia1010021

Skavronskaya, L., Moyle, B., & Scott, N. (2020). The experience of novelty and the novelty of experience. *Frontiers in Psychology*, *11*, 322. https://doi.org/10.3389/fpsyg.2020.00322

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine*, *166*(10), 1092–1097. https://doi.org/10.1001/archinte.166.10.1092

Stiles, W. B., Honos-Webb, L., & Surko, M. (1998). Responsiveness in psychotherapy. *Clinical Psychology: Science and Practice*, *5*(4), 439–458. https://doi.org/10.1111/j.1468-2850.1998.tb00166.x

Ta, V., Griffith, C., Boatfield, C., Wang, X., Civitello, M., Bader, H., DeCero, E., & Loggarakis, A. (2020). User experiences of social support from companion chatbots in everyday contexts: Thematic analysis. *Journal of Medical Internet Research*, *22*(3), e16235. https://doi.org/10.2196/16235

Uludag, K. (2023). The use of AI-supported Chatbot in Psychology. *Available at SSRN 4331367*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4331367

Van Dis, E. A., Bollen, J., Zuidema, W., van Rooij, R., & Bockting, C. L. (2023). ChatGPT: Five priorities for research. *Nature*, *614*(7947), 224–226. https://doi.org/10.1038/d41586-023-00288-7

Van Zyl, L. E., Efendic, E., Rothmann, S., & Shankland, R. (2019). Best-practice guidelines for positive psychological intervention research design. In *Positive psychological intervention design and protocols for multi-cultural contexts* (pp. 1–32). Springer International Publishing.

Wagenmakers, E.-J., Verhagen, J., Ly, A., Bakker, M., Lee, M. D., Matzke, D., Rouder, J. N., & Morey, R. D. (2015). A power fallacy. *Behavior Research Methods*, *47*(4), 913–917. https://doi.org/10.3758/s13428-014-0517-4

Waterman, A. S. (2013). The humanistic psychology–positive psychology divide: Contrasts in philosophical foundations. *The American Psychologist*, *68*(3), 124–133. https://doi.org/10.1037/a0032168

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Wu, Y., Wu, W., Xing, C., Zhou, M., & Li, Z. (2016). Sequential matching network: A new architecture for multi-turn response selection in retrieval-based chatbots. *ArXiv Preprint ArXiv:1612.01627*.

Zamfirescu-Pereira, J. D., Wong, R. Y., Hartmann, B., Yang, Q. (2023). Why Johnny can't prompt: How non-AI experts try (and Fail) to design LLM prompts. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (pp. 1–21). https://doi.org/10.1145/3544548.3581388

About the authors

Ivan Liu is a researcher and lecturer at the Department of Psychology, Faculty of Arts and Science, Beijing Normal University at Zhuhai, China. His research focuses on artificial intelligence, large language models, chatbots, affective computing, multimodal emotion recognition, and positive psychology.

Fangyuan Liu is an undergraduate student in the Department of Psychology, Faculty of Arts and Science, Beijing Normal University at Zhuhai, China. Her research interests include multimodal emotion recognition, AI agents, and psychology.

Yuting Xiao is a postgraduate student in the Faculty of Psychology at Beijing Normal University, China. Her research interests include positive psychological intervention, generative chatbots, large language models, and natural language processing.

Yajia Huang, MSc, graduated from the Faculty of Psychology at Beijing Normal University, China. She works as a human factors engineer in the vehicle engineering department for an electric car manufacturer in Shanghai, China.

Shuming Wu is a postgraduate student in the Faculty of Psychology at Beijing Normal University, China. Her research interests include brand marketing, storytelling brand archetypes, natural language processing, and AIGC.

Shiguang Ni is a professor at the Shenzhen International Graduate School, Tsinghua University. His research focuses on the human-computer interface, serious games, multimodal emotion recognition, and computational positive psychology.