

# The investigation of progress and application related to chatbot in psychology

**Zhiran Li**

Psychological Research and Counseling Center, Southwest Jiaotong University,  
Chengdu, 611756, China

[lizhiran@stu.hebmu.edu.cn](mailto:lizhiran@stu.hebmu.edu.cn)

**Abstract.** The discipline of psychology plays a pivotal role in contributing to the development of scientific knowledge and the advancement of society, as well as being a prominent area of scientific research globally. In recent years, an increasingly number of artificial intelligence technologies have been combined with psychology technologies, aiming to facilitate in promoting the development of people's mental health through artificial intelligence technology. Numerous chatbot technologies have emerged over the years, and these chatbots have excellent comprehensive capabilities and have been widely used in the field of psychology. However, there is little literature reviewing the application of chatbots in the field of psychology, and this review aims to summarize the technologies of chatbots and their application in psychology in this last several years, in order to help subsequent scholars understand the relevant research results in this field more quickly. This study summarizes some of the chatbot approaches in recent years and divides them into four parts, which are 1) ChatGPT 2) Chatbot 3) Accessing personality and emotions 4) Mental health support, and this study also further describes the advances, shortcomings, and future perspectives of the research in this field. and future perspectives in this field, hoping to be helpful to subsequent scholars.

**Keywords:** ChatGPT, Chatbots, Psychology, Machine Learning.

## 1. Introduction

Psychology is the scholarly discipline concerned with comprehending the intricacies of individual psychology and behavior, aiming to elucidate the occurrence, development, and functioning of cognitive processes such as attention, perception and functioning of complex psychological phenomena such as personality, intelligence, emotion and motivation. The field of psychology assumes a pivotal role in the advancement of scientific knowledge and societal progress, while concurrently serving as a prominent area of interest in the global landscape of scientific research.

In recent times, the expeditious advancement of Artificial Intelligence (AI) has engendered increasingly sophisticated chat robot and Large Language Models (LLMs), such as OpenAI's ChatGPT. ChatGPT is an AI-based conversational agent that utilizes natural language processing (NLP) and machine learning algorithms to simulate human-like conversations [1]. ChatGpt has a huge market potential and has received a hot response from users upon release [2]. An increasing number of scholars have extensively investigated the potential application scenarios of ChatGPT, revealing its significant promise within the domain of psychology. First of all, ChatGPT has been demonstrated to

generate creative ideas [3], while it outperforms humans in emotional awareness evaluations [4]. There are also some evidences show that ChatGPT performance similar to those of an average human subject on verbal insight problems [5]. All these outstanding abilities show how ChatGPT is different from previous text production tools. Secondly, ChatGPT contributes a lot to the field of mental health, it can help to assess human personalities [6], ChatGPT can also detect depression [7], and distinguish paranoid thoughts in patients with Schizophrenia [8]. More importantly, ChatGPT can provide appropriate psychological assistance to specific populations, such as assisting children with Down Syndrome [9], and support pediatric behavioral health in a number of ways [10]. Psychological support and advice for anxiety disorders [11], paratherapist counseling [12] as well as psychological services to address sexual violence [13]. In addition to this, ChatGPT can also provide treatment for many psychological disorders or symptoms, for example it could provide personalized obesity treatment [14]. In addition to ChatGPT's extraordinary capabilities and contributions to the field of mental health as mentioned above, it can also be used as an effective tool for human-computer interaction research, helping to analyze textual data [15].

However, despite the fact that scholars have done a lot of research and discussion on the application of ChatGPT in the field of psychology, there are still no studies that have done review articles in this area. To address this gap, the present paper aims to consolidate the contributions of ChatGPT to the field of psychology over recent years. By doing so, it seeks to facilitate subsequent scholars in efficiently navigating the extant research within this domain and further foster the prosperous advancement of ChatGPT's application in psychology. This study will focus on the methods behind different chatbots and their wide application in the field of psychology, focusing on the following key aspects: (i) ChatGPT, (ii) Chatbot, (iii) Assessing personality and emotions, (iv) Mental health support. These results do more than just add to the existing literature, taking the benefits of chatbots in psychology a step further. At the same time, this study also discloses some dilemmas of chatbots, which hopefully future scholars can build on to address these issues with further research.

## 2. Method

### 2.1. ChatGPT

*2.1.1. The underlying technology of ChatGPT.* ChatGPT, a chatbot program developed by OpenAI, was released on November 30, 2022. The underlying technology of ChatGPT is mainly based on the Transformer model, which is a sequence-to-sequence model using the self-attention mechanism, proposed by Google in 2017. The architecture of the Transformer model is highly parallelizable, mainly consisting of the following parts: Tokenization embedding (Embedding), positional encoding, Transformer module, Softmax function etc.

*2.1.2. Training.* ChatGPT training is divided into two steps: pre-training and fine-tuning. 1) Pre-training. In this phase, the model is exposed to a large amount of textual data from the Internet. However, it is important to note that the model does not know the details of which documents are in its training set, nor does it have access to any particular document or source. The goal of this phase is to learn the statistical patterns of the language. 2) Fine-tuning. After pre-training, the model undergoes a fine-tuning process where it is trained on a narrower dataset with the help of a human reviewer, following certain guidelines provided by OpenAI. This dataset consists of various hypothetical inputs and responses. This phase helps the model to respond better to specific inputs and maintain a safer and more useful interaction with the user. 3) Generating Responses. After training, ChatGPT generates responses using a method called "autoregression". It starts with the input message and then predicts the next word and the next until a complete sentence is formed. It selects each word based on the predicted probability of the next possible word and considers all previous words in the input. This process continues until specific conditions are met, such as reaching the maximum number of words.

ChatGPT sometimes introduces randomness into the word selection process to produce diverse and creative responses.

## 2.2. Chatbot

During epidemics, individuals face significant health threats, and there is a notable increase in mental health issues like depression and anxiety [16]. As mental disorders have been on the rise in recent years, the importance of mental health has become a widespread concern in today's society. The advancements in artificial intelligence and natural language processing have brought attention to the potential of machines as human-like conversational agents. Recent studies have focused on developing chatbots based on Cognitive Behavioral Therapy (CBT) principles, providing cognitive support to users. These interventions have proven effective in reducing psychological distress and alleviating the burden on individuals, particularly for depression and anxiety [17]. This study will present Emohaa, a hybrid system that includes a platform based on the principles of CBT and cognitive support, as well as a conversational platform for emotional support covering a variety of topics. This hybrid system effectively helps people with mental illnesses, aiming to provide them with emotional help and offer them some specialized advice and guidance.

*2.2.1. Emohaa.* The study introduces a dialog agent that comprises of two platforms. The initial platform is based on templates and incorporates pre-defined choices and practiced conversations. This platform assists participants in enhancing their psychological well-being by following the principles of cognitive behavioral therapy (CBT). The second platform, on the other hand, is a generative dialog system that enables users to engage in free-flowing discussions regarding various emotional concerns. Ultimately, this platform offers emotional support to users.

*2.2.2. Based on CBT.* Based on the principles of counseling therapy, the study integrated two different practices: automatic thought training and guided expressive writing. The study utilized more than 20 guided expressive writing exercises in which a variety of topics were covered, and users were guided throughout the process.

*2.2.3. ES-based.* This study utilized Liu et al.'s Emotionally Supportive Conversation dataset (ESConv) [18] and translated the dialog into Chinese. In this study, when building an ES-based platform, a policy-controlled emotion-supported dialog model based on a large Chinese dialog model was built. The study also trained an additional model [19] to categorize whether a user's messages exhibit suicidal thoughts or not, to ensure user safety. Thus, when the appropriate signs are detected, the platform recommends the contact information of the relevant organizations.

*2.2.4. Materials.* To evaluate the efficacy of chatbots in improving mental health, a total of 301 participants were enrolled in this study and assigned to three groups: Emohaa (CBT-based), Emohaa (full), and the control group. As previously mentioned, the ES-based platform enabled users to input their desired text messages and generate responses using a generative model. On the other hand, the CBT-based platform involved template-based conversations. Participants in the Emohaa (full) group had access to both platforms. The subjects' depression condition, anxiety level, mood, and insomnia severity were assessed through pre- and post-testing.

## 2.3. Assessing personality and emotions

Recent research suggests that large language models may have human-like features of self-improvement and reasoning [20]. In a growing body of research, it has been recognized that large prophetic models have virtual personalities and psyches, and that these traits of large language models play an important role in guiding their response and interaction patterns [21]. Zohar Elyoseph tested ChatGpt's ability to recognize emotions, and his study utilized the Levels of Emotional Awareness Scale (LEAS) as an objective, performance-based performance-based test, he analyzed ChatGPT's

responses to 20 scenarios and tested ChatGPT's emotionally aware performance in each scenario versus that of an average person and reviewed the performance of emotionally aware persons, i.e., ChatGPT's emotionally aware performance versus that of an average person. A second examination was conducted after the second month to measure improvement in emotional awareness over time. Finally, two independent licensed psychologists assessed the fit of the ChatGPT's emotionally aware responses to the situation [5]. Based on this assumption that ChatGPT have a virtual personality and psyche, several works [21-23] have applied psychological tests such as the Big Five factors [24] to assess their pseudo-personality (e.g., behavioral tendencies) and thus detect social and ethical risks (e.g., racial bias) in their applications. The present study introduces the novel idea of having the Big Language Model assess human personality and proposes a generalized assessment framework for obtaining quantitative assessments of human personality (e.g., personality types and dispositions) from the Big Predictive Model via the Myers-Briggs Type Indicator (MBTI) [25]. Through these studies, the big language model is able to assess different personalities and emotions, which will have a wider application in different scenarios in the future.

#### *2.4. Mental health support*

The issue of mental health has always been a central concern for humanity. In recent years, the rapid advancement of Natural Language Processing (NLP) technology, as evidenced by ChatGPT and GPT-4, has revolutionized conversational AI capabilities [26-28]. Large-scale language modeling (LLM) has proven successful in various natural language understanding tasks and has potential applications in mental health research based on NLP. Huachuan Qiu introduces the SMILE approach, an inclusive language extension technique that leverages ChatGPT to extend single-turn conversations into multi-turn conversations. This study conducted a comprehensive and systematic comparative analysis of datasets generated with and without the use of the SMILE method. The final findings demonstrate that employing the SMILE methodology results in the generation of a large, diverse, and highly realistic corpus of multi-turn mental health support conversations. This corpus encompasses vocabulary, semantic features, and conversation topics. Consequently, in this study, researchers utilized the collected corpus (SMIECHAT) to develop a more effective dialogue system. The objective is to provide emotional support and constructive advice to a greater number of individuals suffering from mental illnesses.

### **3. Applications and discussion**

#### *3.1. Chatbot*

Recent work in this field has focused on providing cognitive support through conversational agents that employ the principles of CBT. It is expected that there will be an increasing number of studies investigating the impact of chatbots on psychology. Although chatbots have a wide range of applications in psychology, chatbots such as ChatGPT cannot be conscious and have memories as humans do, ChatGPT also tends to produce overconfident responses, even though they may be wrong or inaccurate. And it is not reliable enough to handle sensitive or inappropriate content and may generate inappropriate responses. Furthermore, ChatGPT cannot store personalized information and thus give different answers to different individuals. Chatbots may sometimes still have some problems with sentence breaks and recognition, and in the future, the correctness of statement recognition and the accuracy of responses can be improved. Future research can further investigate the use of chatbots in psychology through empirical design. As well as improving the interface of the chatbot on a user experience level and considering adding more features in the future. For example, the introduction of vivid visualization of chatbot-approved products can be also considered [29].

#### *3.2. Assessing personality and emotions*

As one of the basic individual characteristics, personality describes the relatively stable pattern of individual w.r.t. her/his behavior, thought, and emotion [30]. In recent years, more and more

researchers consider personality as a valuable factor and incorporate it into various tasks to significantly improve performance. In order to automatically obtain large-scale user personalities, text-based personality recognition tasks have been designed to infer user personalities from given user-generated texts [31]. With the rapid development of pre-trained Large Language Models (LLMs), more and more LLM-based models have been proposed for text-based personality detection tasks with significant performance improvements [32]. However, ChatGPT exhibits unfairness to some sensitive demographic attributes, leading to unfair treatment of some specific groups when predicting personality. Although ChatGPT has a wide range of applications in different fields, its scientific validity is also still being explored. As for future work, on the one hand, it would like to apply level-oriented prompting strategy to more NLP tasks for observing its effectiveness in mining text information. On the other hand, with the continuous emergence of various LLMs, future work are interested in exploring the construction of domain-specific LLMs based on psychological data in order to enhance the personality recognition ability of LLMs.

### *3.3. Mental health support*

In recent years, the increasing use of artificial intelligence technology has led to the widespread application of dialog platforms in the field of mental health. ChatGPT, for instance, not only helps identify symptoms of mental illnesses but also offers psychological support and assistance to individuals with different mental health conditions. However, these dialog platforms also come with certain limitations. Firstly, individuals using AI-based apps for mental health purposes are required to share personal data, making them vulnerable to potential leaks and data breaches. Secondly, the lack of proper standardization and monitoring has resulted in the prevalence of apps that may lead to misdiagnosis, incorrect diagnosis, inappropriate recommendations, and an inability to handle crises. These ethical concerns related to the use of ChatGPT and AI-based applications encompass academia, diagnosis, treatment, and therapy. Additionally, Adlung et al. have identified two challenges, namely interpretability and causality, faced by AI in clinical decision making. Interpretability refers to the model's ability to provide results such as statistically significant factors associated with the output. While ChatGPT may explain its answers or reasons for not having the correct answer based on examples found on the web, it does not necessarily mean that ChatGPT can provide plausible explanations. There is still potential for language models to improve transparency by facilitating explanations and enhancing the understanding of deep learning models.

## **4. Conclusion**

This study provides a comprehensive overview of the application of different dialog platforms in psychology. Through the integration, it can be found that these conversation platforms have a high level of competence, mainly in the form of high level of emotion recognition, ability to generate ideas, cognitive decision-making, and have a wide range of applications in the field of mental health aimed at people with mental illnesses. Also, ChatGPT has applications in the field of human-computer interaction. However, this study only focused on the use of chat robot in psychology and did not discuss the use of other AI technologies in psychology. Also, the number of literatures searched in this study is limited. Future research should retrieve more literature while focusing more attention on the application of other AI technologies in the field of psychology.

## **References**

- [1] Hassani H et al 2023 The Role of ChatGPT in Data Science: How AI-Assisted Conversational Interfaces Are Revolutionizing the Field. Big Data and Cogn. Comput
- [2] Hu Krystal 2023 ChatGPT sets record for fastest-growing user base – analyst note Reuters Archived from the original on February 3 2023 Retrieved June 3
- [3] Uludag K 2023 Testing creativity of ChatGPT in psychology: Interview with ChatGPT Available at SSRN 4390872

- [4] Elyoseph Z Hadar-Shoval D Asraf K et al 2023 ChatGPT outperforms humans in emotional awareness evaluations *Frontiers in Psychology* 14: 1199058.
- [5] Orrù G Piarulli A Conversano C et al 2023 Human-like problem-solving abilities in large language models using ChatGPT *Frontiers in Artificial Intelligence* 6: 1199350
- [6] Rao H Leung C Miao C 2023 Can chatgpt assess human personalities? a general evaluation framework *arXiv preprint arXiv:2303.01248*
- [7] Lamichhane B 2023 Evaluation of chatgpt for nlp-based mental health applications *arXiv preprint arXiv:2303.15727*
- [8] Uludag K 2023 ChatGPT Can Distinguish Paranoid Thoughts in Patients with Schizophrenia Available at SSRN 4391941
- [9] Yang J Por L Y Leong M C et al 2023 The Potential of ChatGPT in Assisting Children with Down Syndrome *Annals of Biomedical Engineering* 1-3
- [10] van Schalkwyk G 2023 Artificial intelligence in pediatric behavioral health *Child and Adolescent Psychiatry and Mental Health* 17(1): 1-2
- [11] Farhat F 2023 ChatGPT as a complementary mental health resource: a boon or a bane *Annals of Biomedical Engineering* 1-4
- [12] Carlbring P Hadjistavropoulos H Kleiboer A et al 2023 A new era in Internet interventions: The advent of Chat-GPT and AI-assisted therapist guidance *Internet Interventions* 32
- [13] Folastrì S Hambali I M Ramli M et al 2023 ChatGPT educates college students about sexual violence and its impact on their mental health: a proposal *Journal of Public Health* fdad078
- [14] Arslan S 2023 Exploring the Potential of Chat GPT in Personalized Obesity Treatment. *Annals of Biomedical Engineering* 1-2
- [15] Tabone W and De W J 2023 Using ChatGPT for human–computer interaction research: A primer *Manuscript submitted for publication.*
- [16] Lakhan R Agrawal A Sharma M 2020 Prevalence of depression, anxiety, and stress during covid-19 pandemic *Journal of Neurosciences in Rural Practice* 11
- [17] 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. doi:10.2196/mental.7785.
- [18] Liu S et al 2021 Towards emotional support dialog systems, in: *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, Association for Computational Linguistics, Online. pp. 3469–3483. URL: <https://aclanthology.org/2021.acl-long.269>
- [19] Liu Y et al 2019 Roberta: A robustly optimized BERT pretraining approach *CoRR* abs/1907.11692. URL: <http://arxiv.org/abs/1907.11692>, arXiv:1907.11692.
- [20] Huang J X 2022 Large language models can self-improve. *arXiv preprint arXiv:2210.11610*.
- [21] Jiang G Y et al 2022 MPI: Evaluating and inducing personality in pre-trained language models *arXiv preprint arXiv:2206.07550*
- [22] Li X X et al 2022 Is GPT-3 a psychopath? evaluating large language models from a psychological perspective *arXiv preprint arXiv:2212.10529*
- [23] Karra S R et al 2022 AI personification: Estimating the personality of language models *arXiv preprint arXiv:2204.12000*
- [24] Digman J M 1990 Personality structure: Emergence of the five-factor model. *Annual review of psychology* 41(1):417–440
- [25] Myers I B et al 1985 *Manual: A guide to the development and use of the Myers-Briggs Type Indicator* Consulting Psychologists Press
- [26] Vaswani A et al 2017 Attention is all you need *Advances in neural information processing systems* 30

- [27] Long O Y et al 2022 Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, 35:27730–27744.
- [28] Ni J J et al 2022 Recent advances in deep learning based dialogue systems: A systematic survey *Artificial intelligence review* pages 1–101
- [29] Jin S V & Youn S 2022 Social Presence and Imagery Processing as Predictors of Chatbot Continuance Intention in Human-AI-Interaction *International Journal of Human–Computer Interaction* 1-13
- [30] Edward D and Richard E L 2019 Personality traits. *General psychology: Required reading* 278
- [31] Hao L et al 2023 A novel personality detection method based on high-dimensional psycholinguistic features and improved distributed gray wolf optimizer for feature selection. *Information Processing & Management*, 60(2):103217
- [32] Dipika J et al 2022 Personality bert: A transformer-based model for personality detection from textual data. In *International Conference on Computing and Communication Networks*, pages 515–522