Applications of AI-powered Conversational Chatbot for Mental Health

Yiyi Chang

Faculty of Art, University of British Columbia, Vancouver, British Columbia, V6T 1Z4, Canada

yiyich@students.cs.ubc.ca

Abstract. Psychological issues have become a pervasive problem affecting the quality of life for many individuals. However, the scarcity of professional psychotherapists and the high threshold for receiving human psychological counselling prevent most people from obtaining timely, high-quality psychological help. To address this issue, chatbots have been developed to participate in the promotion of public mental health. Empowered by Artificial Intelligence (AI), especially Large Language Models (LLMs), chatbots have the potential to revolutionize the field of mental health by offering personalized and full-time support. Moreover, AI-powered chatbots can assist researchers in collecting more data to understand mental health better and develop more effective treatments. This paper categorizes and summarizes the recent applications of conversational chatbot technology in the mental health field, including human-robot relationships, the use of conversational chatbots with mental health tasks in counselling and online settings, the generation of counselling dialogue data, and the evaluation of datasets and models. The advantages and disadvantages of these technologies are explored, along with the current technical shortcomings of conversational chatbots. Additionally, the challenges to their widespread adoption and use, as well as future directions for development, are discussed.

Keywords: Conversational chatbot, mental health, artificial intelligence.

1. Introduction

While the human body is not designed to cope with long-term social stress, the development of society, coupled with the popularization of modern psychology, has led to an increasing awareness of the impact of psychological issues on health and quality of life. According to a World Health Organization (WHO) report, in 2019, approximately 970 million people were battling psychological problems, with anxiety and depression being the most common symptoms [1]. This number went through an acute increase during the subsequent pandemic, rising by about 27% within a year [1]. From mild symptoms causing negative emotions to moderate and severe symptoms leading to somatization and even suicidal attempts, untreated anxiety and depression can significantly reduce quality of life. Despite the large number of people in need of psychological support, only a small proportion receive effective treatment. This disparity is primarily due to the high costs and time required to train medical professionals, resulting in a shortage of qualified personnel. Concerns over market quality inconsistency, privacy issues, and personal preferences further limit access to proper psychological care.

Since the release of Generative Pre-trained Transformer (ChatGPT) by OpenAI in 2022, which demonstrated remarkable capabilities in generating human-like dialogues, an increasing number of scientists have begun exploring how conversational chatbots, programs designed to simulate human conversation, can bridge the gap in public mental health care left by human psychological counsellors [2]. However, the interest in chatbots began nearly 80 years earlier, in the 1960s, when Joseph Weizenbaum developed the first chatbot ELIZA [3]. ELIZA was a rule-based chatbot designed to simulate a psychotherapist's conversation [3]. It did not understand the conversation, but generated responses based on predetermined templates and user inputs [3]. Despite its technical simplicity and inability to engage in deep conversations, ELIZA sparked scientists' interest in chatbots. Over time, with the development of Artificial Intelligence Markup Language (AIML), deep learning, and Natural Language Processing (NLP) technologies, chatbots have become increasingly capable of contributing to public mental health promotion in various ways.

In this paper, recent applications of conversational chatbot technology in mental health are reviewed and categorized. The major categories include the potential of conversational chatbots to impact mental health, the application of chatbots integrated with psychological knowledge in various domains, the ability of generative language models to expand counselling databases, and the evaluation of datasets and models. The advantages and disadvantages of these technologies are explored, along with the current technical shortcomings of conversational chatbots. Additionally, the challenges to their widespread adoption and use, as well as future directions for development, are discussed.

2. AI-powered Chatbots

2.1. Conversational Chatbot and Human Relationships

A major direction in the development of conversational models is to exhibit more human-like behaviours, thus fostering trust-based relationships with users. In a historical review of chatbots, Go and Sundar identified factors that could make chatbots appear more human-like, including appearance, naming, and contextual continuity in conversations [4]. They categorized these factors into three main types: anthropomorphic visual presentation, identity cues, and conversational cues [4]. Their experimental validation revealed that the linguistic abilities of chatbots to construct deep, flexible dialogues through contextual interaction most significantly influence users' psychology, attitudes, and behaviours during interactions [4].

Just like in human-to-human relationships, people have their preferences for chatbots with different appearances and abilities, and they can establish connections with chatbots through interaction. Skjuve et al. explored how individuals establish emotional connections with chatbots [5]. Guided by the Social Penetration Theory, which explains human relationships, they examined how users build trust with the social chatbot Replika through one-way self-disclosure and in-depth conversations [5]. The way trust is built between users and chatbots is similar to the relationship between patients and therapists in psychological counselling. Skjuve's work demonstrates the feasibility of using online psychological counselling chatbots as a substitute for human counsellors.

2.2. Conversational Chatbot in Counseling

In recent years, conversational chatbots have assumed significant roles in people's daily lives, such as in customer service, educational tutoring, and personal assistance. In advance, there have been attempts to utilize conversational chatbots in professional fields. An undifferentiated large language model (LLM) has demonstrated its ability to pass multiple professional medical tests, showcasing the potential for these language models to be used as assistants in the medical field to enhance work efficiency [6]. Hu et al. also demonstrated that, while performance depends on task complexity, GPT shows potential in diagnostic applications [7].

In the field of psychology, there is extensive research on using conversational chatbots to promote mental health, such as assisting with counselling processes. Leveraging the dialogue simulation capabilities of language models, the ChatGPT Client Simulation (CCS), a model based on ChatGPT, was developed to help counsellors train counselling skills through role-play [8]. GPT has also been used to assist traditional counselling [9]. Although it is not recommended for providing medical advice, GPT can serve as a professional personal assistant to therapists by offering therapy suggestions, providing emotional support between sessions, or recording key points and client information during conversations to reduce the therapist's workload [9, 10]. If specific requirement is needed, Omarov et al. demonstrated applications of language models combined with particular counselling approaches, such as cognitive-behavioural therapy (CBT) [11]. Beyond trained counsellors, many untrained individuals encounter situations requiring psychological intervention on social platforms, such as when someone expresses strong suicidal tendencies [12]. Fu et al. improved an online helping platform based on GPT and the LLM-Counselors Support System to assist non-

professional volunteers and clients in enhancing conversation quality and efficiency, greatly increasing the availability and quality of psychological helping resources [12].

There is also work on using fine-tuned large language models for standalone online psychological counselling. Psy-LLM, designed by Tin et al., is a question-answering framework aimed at providing online counselling services [13]. LLMs rely on databases for their operations. In this regard, Tin's team utilized multiple datasets focused on psychological Question and Answer (Q&A), including PsyQA, an evaluated dataset widely used in AI-based counselling chatbot research, containing 22K questions and 56K well-structured answers [14], as well as other datasets collected from social media platforms that include Q&A form posts about psychological issues, and evaluated by the team [13]. They trained two pre-trained LLMs, PanGu and WenZhong, using these datasets, testing the scores of their output in helpfulness, fluency, relevance, and logic. Ultimately, PanGu was selected as their base model due to its higher overall scores [13]. It is noteworthy that both models did not achieve the average human rating for ground-truths from the datasets, indicating that the technique of fine-tuning pre-trained LLMs with psychology-specific databases still requires further technical upgrades and follow-up [13].

2.3. Conversational Chatbot to Generate Dialogue Dataset

When training conversational chatbots to simulate or substitute real psychological counselling, datasets serve as crucial resources for fine-tuning the model's responses. However, due to confidentiality and privacy concerns inherent in psychological treatment, the number of available professional databases is insufficient for the models to achieve highly human-like performance. To address this, language models are not only used as the basis for counselling models but also employed to generate psychological counselling dialogues for training databases. SMILE (Single-turn to Multi-turn Inclusive Language Expansion), introduced by Qiu et al., leverages ChatGPT to convert single-turn dialogues into multi-turn exchanges, simulating real counselling processes [15]. This approach offers a novel method for generating diverse dialogue datasets [15]. Zhang et al. created a process to transform psychological counselling reports into multi-turn consultation dialogues [16]. Their collected reports, after privacy shadowing, undergo the Memo2Demo dialogue reconstruction method, a two-phase pipeline involving a psychological supervisor and a counsellor [16]. In this method, the supervisor first converts the raw report into a counselling note that includes basic information about counselling and an elaborate consultation plan. The counsellor then refines the report into a multi-turn dialogue based on the raw report and a frame that simulates the actual psychological counselling process [16]. Both studies aim to address the current deficiency of consultation data in training datasets and offer innovative approaches for generating simulated real-world data.

2.4. Evaluation of Datasets and Models

For conversational models intended for psychological counselling, ensuring their professionalism is critical to protect the user's mental well-being and avoid causing further harm. Comprehensive evaluation of these models is thus a necessary task. Different evaluation metrics have been employed to test the datasets used for training models [16]. In Zhang et al.'s work, they evaluated the transformed dialogues on comprehensiveness, professionalism, authenticity, and safety [16]. Qiu et al. used multiple existing evaluation metrics, such as Perplexity (PPL) for model prediction accuracy, Metric for Evaluation of Translation with Explicit Ordering (METEOR) for word order and morphological matching, Bilingual Evaluation Understudy (BLEU)-1/2/3 for translation quality based on n-gram matching, Rouge-L for the longest common subsequence between candidate and reference texts, and Distinct-1/2/3 (D-1/2/3) for generated text diversity [15]. Both research groups emphasized the importance of testing textual performance and logical consistency while also employing psychological experts to manually evaluate their datasets.

Another significant development is the PsyEval, an evaluating process model constructed by Jin et al. and is the first model that comprehensively evaluates LLMs with mental health-related tasks [17]. PsyEval tests language models in three major areas: knowledge, diagnosis, and emotional

support [17]. Specifically, this evaluation model expects language models to demonstrate (1) the ability to pass basic mental health-related knowledge and major mental health crisis Q&A tests, (2) the capacity to diagnose various psychological disorders using online text data and complex patient-doctor dialogue records, and (3) being able to provide emotional support such as empathy and safety during counselling sessions [17]. PsyEval provides a comprehensive approach to testing the capabilities of conversational models. Although it covers a wide range of tests, detailed, in-depth evaluation in each area is still needed for precise results, especially in the evaluation dialogue quality.

3. Discussion

3.1. Benefits and Drawbacks of Using Conversational Models

Conversational models offer significant advantages in promoting public mental health, being able to support professional counsellors' work and providing immediate assistance to non-professionals. The capability of large language models to function independently as online counselling platforms significantly expands access to professional, reliable mental health care resources [12]. Additionally, generative chatbots can mitigate the lack of diversity in training datasets by generating simulated dialogues through their abilities in emotional and situational analysis [15, 16].

However, the potential drawbacks of using conversational models for counselling services have not been thoroughly discussed. Research evaluating the effectiveness of these models often lacks long-term follow-up and a persuasive number of participants due to privacy, user attrition, and funding issues. As a result, the impact of online models on public mental health is not conclusively proven. Many studies focus on whether these models can alleviate common psychological symptoms like depression and anxiety, but there is limited research on whether they might negatively affect users' mental health and daily lives. Laestadius et al. highlight that users may suffer harm from limited access and inappropriate responses when they become emotionally dependent on social chatbots [18]. To gain wider acceptance and use in the future, the effectiveness and potential drawbacks of these counselling models need further evaluation.

3.2. Technical Limitations

Several technical issues need to be addressed when using conversational models for public mental health, such as the high similarity among different online counselling models, the extent of fine-tuning based on specific tasks, model and dataset evaluation, personalization settings, and considerations for safety and privacy. Most existing online counselling models are constructed under a similar mechanism, combining pre-trained LLMs and counselling dialogue datasets such as ChatGPT and PsyQA [14]. The difficulty in obtaining professional counselling dialogue data and the limited quantity of such data result in these models having little difference in performance and characteristics. Future research could focus on using generative language models to obtain more diverse and realistic dialogue data.

Fine-tuning levels also need exploration. Jin et al. found that models fine-tuned for specific tasks often show higher professional levels but reduced language expression capabilities [17]. In addition to balancing the model's professional and linguistic proficiency through fine-tuning adjustments, research focusing on constructing comprehensive large language models capable of rapidly switching between capabilities in multiple domains, through techniques such as multi-model ensembles, multi-mode models, and meta-learning, can be further explored in the future.

The models with mental health-related tasks and their training datasets must undergo strict evaluation due to their direct impact on users' mental states. Apart from a few widely used, professionally verified datasets like PsyQA, different research groups often use web crawlers to gather data from social platforms and forums, which may lack stringent professional regulation. These datasets are processed and evaluated by various aspects of the data, without a unified standard. The professional and effective evaluation of model outputs remains a research gap as well. Establishing

comprehensive and consensus-driven evaluation standards for databases and models is essential for improving counselling models' safety and dissemination potential.

Societal demands and considerations are keys to enhance the application of conversational models in real life. Due to individual differences and changing times, a universal counselling model is not likely to suit everyone in the long term. Dosovitsky and Eduardo note that young people's perceptions of using emojis in communication changed completely over a few years, indicating the rapid transformation of social media pop culture [19]. However, few existing models support personalized settings or undergo large-scale updates according to pop culture trends. Counselling models might also vary in effectiveness depending on the specific psychological issues they address, yet research results in this area have not been consistent. Klos et al.'s study did not replicate the findings of previous studies and found no significant results for the social chatbot Tess in reducing depression and anxiety symptoms among university students [20-22]. Additionally, these studies on whether conversational chatbots can alleviate users' anxiety or depression symptoms primarily involved young female college students, leaving a gap in research on the effects on different genders, age groups, and psychological issues. Despite attention to data privacy during collection and online usage phases [11], public trust in online counselling platforms' privacy remains low due to data transparency concerns and unfamiliarity with internal model mechanisms. These gaps in personalization and privacy considerations may explain why online counselling models have not gained widespread adoption despite the severe shortage of mental health resources.

4. Conclusion

Despite the significant potential of conversational chatbots in enhancing public mental health, further regulation and research are necessary to ensure responsible implementation. For research groups focusing on evaluating the impact of language models on user symptoms, more long-term, credible follow-up studies are essential to reach definitive conclusions. Moreover, the potential adverse effects of using language models to mitigate mental health symptoms need to be thoroughly explored to prevent negative consequences from widespread adoption.

To improve the practicality of conversational chatbots, several technical shortcomings must be addressed. Researchers should focus on developing more diverse counselling databases, constructing larger, scenario-adaptive combined chatbots, and creating personalized user settings to mitigate the issue of high overlap among different models.

Public acceptance of chatbots is also crucial. To build trust, the industry needs a unified, high-standard testing framework for evaluating the safety and securing privacy of databases and language models. Legal regulations should limit the types of user data that chatbots can collect and clarify data usage, presenting this information in an accessible manner to the public.

References

[1] World Health Organization. Depressive Disorder (Depression). URL: www.who.int/news-room/fact-sheets/detail/mental-disorders. Last Accessed: 2024/08/07.

[2] Lin Chien-Chang, Anna YQ Huang, Stephen JH Yang. A review of ai-driven conversational chatbots implementation methodologies and challenges (1999–2022). Sustainability, 2023, 15 (5): 4012.

[3] Lo Andrew W, Singh Manish, Musumeci Jim, et al. From ELIZA to ChatGPT: The Evolution of Natural Language Processing and Financial Applications. The Journal of Portfolio Management, 2023, 49 (7): 201-235.

[4] Go Eun, S. Shyam Sundar. Humanizing chatbots: The effects of visual, identity and conversational cues on humanness perceptions. Computers in human behavior, 2019, 97: 304-316.

[5] Skjuve Marita, Folstad Asbjorn, Fostervold Knut Inge, et al. My chatbot companion-a study of human-chatbot relationships. International Journal of Human-Computer Studies, 2021, 149: 102601.

[6] Kung Tiffany H, Cheatham Morgan, Medenilla Arielle, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. PLoS digital health, 2023, 2 (2): e0000198.

[7] Hu Yanni, Hu Ziyang, Liu Wenjing, et al. Exploring the potential of ChatGPT as an adjunct for generating diagnosis based on chief complaint and cone beam CT radiologic findings. BMC Medical Informatics and Decision Making, 2024, 24 (1): 55.

[8] Maurya Rakesh K. Using AI Based Chatbot ChatGPT for Practicing Counseling Skills through Role-play. Journal of Creativity in Mental Health, 2023: 1-16.

[9] Eshghie Mahshid, Mojtaba Eshghie. ChatGPT as a therapist assistant: a suitability study. arXiv preprint, 2023: 2304.09873.

[10] Adhikary Prottay Kumar, Srivastava Aseem, Kumar Shivani, et al. Exploring the Efficacy of Large Language Models in Summarizing Mental Health Counseling Sessions: Benchmark Study. JMIR Mental Health, 2024, 11: e57306.

[11] Omarov Batyrkhan, Zhumanov Zhandos, Gumar Aidana, et al. Artificial intelligence enabled mobile chatbot psychologist using AIML and cognitive behavioral therapy. International Journal of Advanced Computer Science and Applications, 2023, 14 (6): 137-146.

[12] Fu Guanghui, Zhao Qing, Li Jianqiang, et al. Enhancing psychological counseling with large language model: A multifaceted decision-support system for non-professionals. arXiv preprint, 2023: 2308.15192.

[13] Lai Tin, Shi Yukun, Du Zicong, et al. Psy-llm: Scaling up global mental health psychological services with ai-based large language models. arXiv preprint, 2023: 2307.11991.

[14] Sun Hao, Lin Zhenru, Zheng Chujie, et al. Psyqa: A chinese dataset for generating long counseling text for mental health support. arXiv preprint, 2021: 2106.01702.

[15] Qiu Huachuan, He Hongliang, Zhang Shuai et al. Smile: Single-turn to multi-turn inclusive language expansion via chatgpt for mental health support. arXiv preprint, 2023: 2305.00450 (2023).

[16] Zhang Chenhao, Li Renhao, Tan Minghuan, et al. CPsyCoun: A Report-based Multi-turn Dialogue Reconstruction and Evaluation Framework for Chinese Psychological Counseling. arXiv preprint, 2024: 2405.16433.

[17] Jin Haoan, Chen Siyuan, Wu Mengyue, et al. PsyEval: A Comprehensive Large Language Model Evaluation Benchmark for Mental Health. arXiv preprint, 2023: 2311.09189.

[18] Laestadius Linnea, Bishop Andrea, Gonzalez Michael, et al. Too human and not human enough: A grounded theory analysis of mental health harms from emotional dependence on the social chatbot Replika. New Media & Society, 2022: 14614448221142007.

[19] Dosovitsky Gilly, Eduardo Bunge. Development of a chatbot for depression: adolescent perceptions and recommendations. Child and adolescent mental health, 2023, 28 (1): 124-127.

[20] Klos Maria Carolina, Escoredo Milagros, Joerin Angela, et al. Artificial intelligence–based chatbot for anxiety and depression in university students: pilot randomized controlled trial. JMIR formative research, 2021, 5 (8): e20678.

[21] Fitzpatrick Kathleen Kara, Alison Darcy, Molly Vierhile. 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, 2017, 4 (2): e7785.

[22] Fulmer Russell, Joerin Angela, Gentile Breanna, et al. Using psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: randomized controlled trial. JMIR mental health, 2018, 5 (4): e9782.