

# **‘Enabling Empathetic Conversations?’ – A Critical Commentary on AI-Based and Digital Mental Health Interventions**

## **Introduction**

The COVID-19 global pandemic has highlighted significant gaps in mental health services, as corresponding issues including job losses, physical illnesses and mortality, alongside social isolation, led to a stark increase in poor mental health conditions. Recent advances in artificial intelligence (AI) promise significant transformations in healthcare, particularly in relation to digital interventions. According to a report from The World Health Organisation (WHO, 2022b), 970 million people worldwide live with a mental disorder, with anxiety and depression being the most common issues (WHO, 2022a). Furthermore, estimates suggested a 26% and 28% increase, respectively, for anxiety and depressive disorders in just one year (WHO, 2022). Digital mental health apps advocate a myriad of benefits in the domain of self-help, including the convenience of on-demand access to resources, targeting a range of mental health conditions using a single application, and users have the potential to experience better sleep and reductions in stress (Higgins et al., 2023). However, many of these benefits are often made prior to empirical or clinical validation, suggesting that such digital interventions cannot be the only ‘fix’ for the mental health epidemic. Furthermore, recent research has questioned whether these digital mental health apps serve as a quick fix to relieve government pressures on expenditure for preventative and evidence-based treatments (Hamdoun et al., 2023). This blog will present a brief overview of AI-based (e.g., chatbots) mental health interventions in the digital age, critically evaluating whether such interventions enable empathetic conversations for those with profound mental health illnesses.

## **Human-AI collaborative systems**

To understand how AI and digital technologies aid those with mental health issues, it is important to consider the context in which such interactions occur. Prior to AI-based technologies, those with mental health issues have sought peer support via platforms including Reddit and TalkLife (Sharma et al., 2020). Yet, peer supporters face challenge in the realm of empathetic conversations, as they lack appropriate expertise and training in forming strong alliances with support seekers, due to the inability to practise active listening

and relate to sufferers (Sharma et al., 2022). Indeed, *empathetic support* is a critical factor that contributes to strong mental health support, alongside motivational interviewing, and problem-solving skills, which can subsequently lead to symptom improvement and better relationships between the support seeker and peer support (Elliott et al, 2018). Another key issue with online social support is reaching to the masses that require help. To address this issue, introducing human-AI collaborative systems may improve the overall effectiveness of such platforms, by providing automated and actionable responses, with higher levels of empathy. For instance, Sharma and colleagues (2022) conducted a randomised controlled trial, comprising of 300 TalkLife peer supporters who were randomly divided into two groups: **(a)** human only (control) and **(b)** human and AI (treatment) <sup>1</sup>(see **Figure 1 for screenshot of exemplar chat**). Results from the study illustrated that the Human + AI responses were strictly preferred 46.88% of the time relative to a 37.39% strict preference for the Human Only responses. (Sharma et al., 2022).

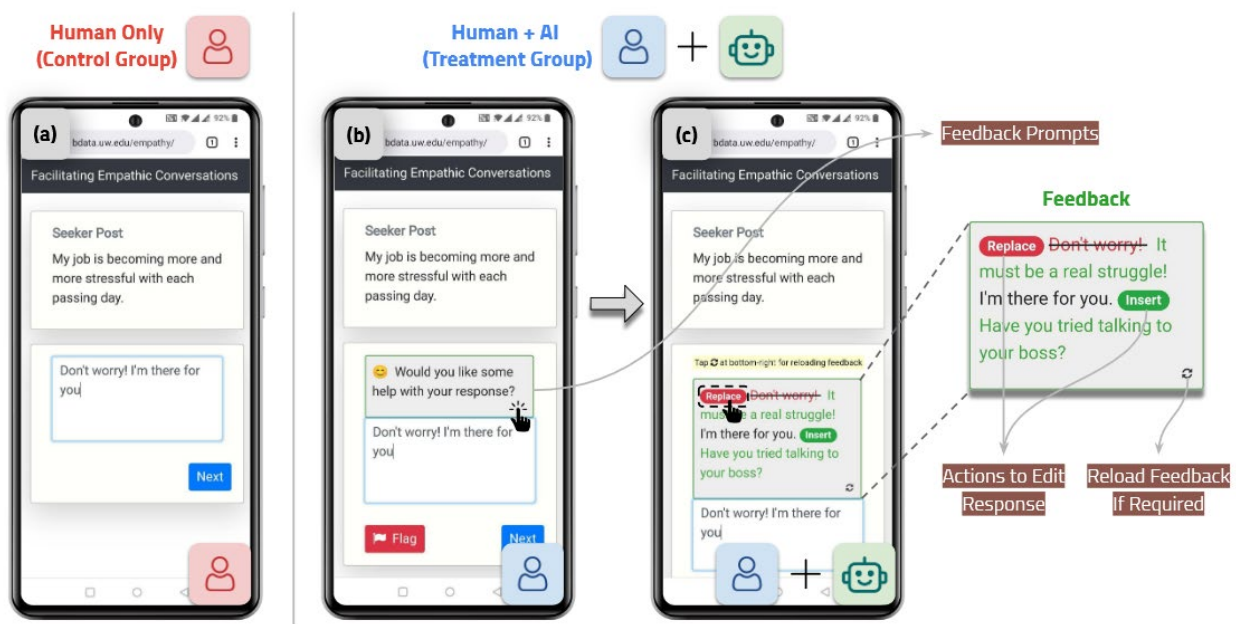


Figure 1. Screenshot of chat taken from Sharma et al., 2022 study

<sup>1</sup> **(a)** Without AI, human peer supporters are presented with an empty chat box to respond to support seekers. **(b)** The feedback agent (HAILEY) prompts peer supporters for providing just-in-time AI feedback as they write their responses, as well as offering changes to make the responses more empathetic.

Despite the promising collaboration between Human-AI systems, there are several limitations to be noted. The measures used to evaluate empathy may not truly capture how the support seeker perceives empathy, and instead collects responses based on empathy that is expressed. Research has suggested that perceived empathy moderates the association between healthy literacy and the understanding of information by patients (Chu & Tseng, 2013). Furthermore, cultural factors may impact the support provided to seekers and should be embedded in future approaches to account for underrepresented minority groups and gender identities. Alongside this, participants were only recruited from TalkLife, and support was only provided in English, thus considering a range of platforms and additional languages may improve the diversity of prospective human-AI systems.

### **AI-chatbots and apps**

AI-chatbots may serve as promising mental health intervention for users, with a range of functions including symptom monitoring, guided meditation, and mood check-ins (Hamdoun et al., 2023). For instance, Sayana was a chat-designed programme aimed at users to increase emotional awareness, by regularly checking in their emotions and associated events, with self-care exercises deeply rooted in dialectical behaviour therapy (DBT), acceptance commitment therapy (ACT), and cognitive behavioural therapy (CBT). To provide a brief overview of each therapy: DBT combines CBT, mindfulness, and behaviourism, and is a type of talking therapy aimed at individuals that experience intense emotions and require necessary skills to manage their emotions more effectively (Bunford et al., 2015). ACT is grounded in mindful psychotherapy, which aims for users to accept their thoughts without judgement and move forward with their emotions, through cognitive diffusion (Harris, 2019). Finally, CBT aims to break vicious cycles of negative thoughts, by working through harmful perceptions and beliefs associated with traumatic experiences (Smith, 2016).

Another chatbot, Youper, has proven popular, being deemed as the most engaging digital mental solution for internalising disorders, such as anxiety and depression, with possibilities to monitor symptoms and to opt for help in various categories including setting goals, reframing thoughts and to stop worrying (Hamdoun et al., 2023). Yet such apps arguably lack the capacity to offer empathy and rationalise reasonably with users, which may lead to people misinterpreting their issues, and becoming increasingly distracted from reality. Another key issue surrounding AI-mental health apps and bots is the following paradox: the promise of being available 24/7 for users, yet there is an expectation for users to be self-sufficient, with the self-management of symptoms. This may be particularly problematic for users seeking emotional support, as reports suggest that AI bots adopt ‘phraseology’ (e.g., “I’m glad I could help!” and “Take care!”), as opposed to human pleasantries, thus rupturing the therapeutic experience.

### **Can we replace human empathy in mental health research?**

The big question at hand is: can AI systems provide emotional empathetic responses to users? Empathy is an umbrella term and can be divided into subcategories: *motivational*, *cognitive* and *emotional empathy*. Whilst emotional and motivational empathy refer to how the experience of emotions leads to empathetic concern from others, cognitive empathy is understood as recognising and detecting specific emotional mental states (Montemayor et al., 2022). Whilst mental health apps and chatbots are a useful way for people to journal their emotional states on a regular basis, it does not divulge into deep rooted issues, particularly when people may experience phobias, are ill or grieving. Thus, the question arises as to whether such interventions should be labelled as appropriate ‘therapeutic tools’, or in fact are they merely ‘advisors’ or just an initial checkpoint for those that have mental health issues? Further research should explore whether perceived and experienced levels of empathy can be addressed in AI-interventions.

Overall, whilst acknowledging the increase in AI algorithms being embedded in mental health research, there is a need to consider their viability and feasibility in real-world applications. Perhaps the promotion of such digital interventions has been augmented, without recognising the key methodological issues, including generalisability to diverse populations and the lack of standardisation across various devices (i.e., not restricting to phones or tablets). Yet, several opportunities are presented in incorporating more AI-based technologies in the mental health sector, including the reduction of hospital burdens, better

preliminary clinical screening of conditions or possible deficits, and an overall improvement in monitoring patients over time. Moreover, integrating diverse communities into such AI-based interventions will be key in future models, accounting for gender identities, racially minoritized backgrounds and sexual orientation. Delving in different forms of empathy and recognising the limitations that AI present in offering appropriate *empathetic support* for users has yet to be researched extensively. This gap should be bridged through further randomised controlled trials, whereby users are offered different levels of empathetic support. Whilst technological advancements are showing promise and potential, it should not supersede the basic infrastructure and necessity for human interactions in the mental health sector.

## References

- Bunford, N., Evans, S. W., & Wymbs, F. (2015). ADHD and emotion dysregulation among children and adolescents. *Clinical Child and Family Psychology Review*, 18(3), 185–217. <https://doi.org/10.1007/s10567-015-0187-5>
- Chu, C.-I., & Tseng, C.-C. A. (2013). A survey of how patient-perceived empathy affects the relationship between health literacy and the understanding of information by orthopedic patients? *BMC Public Health*, 13(1), 155. <https://doi.org/10.1186/1471-2458-13-155>
- Elliott, R., Bohart, A. C., Watson, J. C., & Murphy, D. (2018). Therapist empathy and client outcome: An updated meta-analysis. *Psychotherapy (Chicago, Ill.)*, 55(4), 399–410. <https://doi.org/10.1037/pst0000175>
- Hamdoun, S., Monteleone, R., Bookman, T., & Michael, K. (2023). AI-Based and Digital Mental Health Apps: Balancing Need and Risk. *IEEE Technology and Society Magazine*, 42(1), 25–36. <https://doi.org/10.1109/MTS.2023.3241309>
- Harris, R. (2019). *ACT made simple: An easy-to-read primer on acceptance and commitment therapy*. New Harbinger Publications.
- Higgins, O., Short, B. L., Chalup, S. K., & Wilson, R. L. (2023). Artificial intelligence (AI) and machine learning (ML) based decision support systems in mental health: An integrative review. *International Journal of Mental Health Nursing*, 32(4), 966–978. <https://doi.org/10.1111/inm.13114>
- Montemayor, C., Halpern, J., & Fairweather, A. (2022). In principle obstacles for empathic AI: Why we can't replace human empathy in healthcare. *AI & SOCIETY*, 37(4), 1353–1359. <https://doi.org/10.1007/s00146-021-01230-z>
- Sharma, A., Choudhury, M., Althoff, T., & Sharma, A. (2020). Engagement patterns of peer-to-peer interactions on mental health platforms. In *Proceedings of the international AAAI conference on web and social media*, 14, 614–625. <https://doi.org/10.1609/icwsm.v14i1.7328>
- Sharma, A., Lin, I. W., Miner, A. S., Atkins, D. C., & Althoff, T. (2022). *Human-AI Collaboration Enables More Empathic Conversations in Text-based Peer-to-Peer Mental Health Support*. <http://arxiv.org/abs/2203.15144>
- Singh, O. (2023). Artificial intelligence in the era of ChatGPT - Opportunities and challenges in mental health care. *Indian Journal of Psychiatry*, 65(3), 297. [https://doi.org/10.4103/indianjpsychiatry.indianjpsychiatry\\_112\\_23](https://doi.org/10.4103/indianjpsychiatry.indianjpsychiatry_112_23)
- Smith, A. (2016). A literature review of the therapeutic mechanisms of art therapy for veterans with post-traumatic stress disorder. *International Journal of Art Therapy*, 21(2), 66–74. <https://doi.org/10.1080/17454832.2016.1170055>
- World Health Organisation (2022a) *Mental disorders*. Retrieved from: <https://www.who.int/news-room/fact-sheets/detail/mental-disorders>. Accessed: 20th October 2023.
- World Health Organisation. (2022b). *Mental health and COVID-19: early evidence of the pandemic's impact: scientific brief*, 2 March 2022 (No. WHO/2019-nCoV/Sci\_Brief/Mental\_health/2022.1).

**Biography:** Evelyn Mary-Ann Antony is a PhD researcher in the School of Education at Durham University, and a postgraduate teaching assistant for the module ‘Education, Mental Health and Wellbeing’. Her research focuses on conducting a multidimensional assessment of emotional dysregulation (a key trait that is prevalent in youth psychopathological illnesses) and its longitudinal associations with ADHD and parenting practices, across middle childhood (ages 6-12). She holds a MPhil in Education (Psychology) at the University of Cambridge and a MA (Honours) degree in Psychology at The University of Edinburgh. Please feel welcome to connect with her [here](#) and read about her research [here](#).



**Citation on deposit:**

Antony, E. (2024). 'Enabling Empathetic Conversations?': A Critical Commentary On AI-Based And Digital Mental Health Interventions. [Blog Post - Cambridge Wellbeing & Inclusion SIG]

**For final citation and metadata, visit Durham Research Online URL:**

<https://durham-repository.worktribe.com/output/2516046>

**Copyright statement:** This content can be used for non-commercial, personal study.