Received: 30 April 2022 | Revised: 29 February 2024 | Accepted: 23 April 2024

DOI: 10.1111/bioe.13299

SPECIAL ISSUE: ETHICS OF mHEALTH

bioethics WILEY

Does a lack of emotions make chatbots unfit to be psychotherapists?

Mehrdad Rahsepar Meadi[1,2] | Justin S. Bernstein[3] | Neeltje Batelaan[1] | Anton J. L. M. van Balkom[1] | Suzanne Metselaar[2]

[1]Department of Psychiatry, Amsterdam Public Health, Mental Health program, Amsterdam UMC location Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

[2]Department of Ethics, Law, & Humanities, Amsterdam UMC location Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

[3]Department of Philosophy, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

Correspondence
Mehrdad Rahsepar Meadi, Department of Psychiatry, Amsterdam Public Health, Mental Health Program, Amsterdam UMC location Vrije Universiteit Amsterdam, Boelelaan 1117, Amsterdam, The Netherlands.
Email: m.rahseparmeadi@ggzingeest.nl

Abstract

Mental health chatbots (MHCBs) designed to support individuals in coping with mental health issues are rapidly advancing. Currently, these MHCBs are predominantly used in commercial rather than clinical contexts, but this might change soon. The question is whether this use is ethically desirable. This paper addresses a critical yet understudied concern: assuming that MHCBs cannot have genuine emotions, how this assumption may affect psychotherapy, and consequently the quality of treatment outcomes. We argue that if MHCBs lack emotions, they cannot have genuine (affective) empathy or utilise countertransference. Consequently, this gives reason to worry that MHCBs are (a) more liable to harm and (b) less likely to benefit patients than human therapists. We discuss some responses to this worry and conclude that further empirical research is necessary to determine whether these worries are valid. We conclude that, even if these worries are valid, it does not mean that we should never use MHCBs. By discussing the broader ethical debate on the clinical use of chatbots, we point towards how further research can help us establish ethical boundaries for how we should use mental health chatbots.

KEYWORDS

artificial intelligence, chatbots, countertransference, empathy, mental health, quality of care

1 | INTRODUCTION

Chatbots or 'conversational agents' are computer programmes that mimic human conversation. They are regarded as useful for educational, information retrieval, business, and e-commerce purposes, and their range of application is rapidly expanding.[1] For some years now, chatbots designed to support people with mental health problems have been available.[2] In this paper, we use the term 'mental health chatbot' (MHCB) to refer to such chatbots. These MHCBs are commercially accessible in app stores and operate without direct involvement from mental health professionals. Moreover, this technology is evolving quickly, and chatbots are now being used in ways that resemble traditional psychotherapy.[3] Similar to human therapists, these chatbots are able to engage in

[1]Adamopoulou, E., & Moussiades, L. (2020). An overview of chatbot technology. In I. Maglogiannis, L. Iliadis, & E. Pimenidis (Eds.), *Artificial intelligence applications and innovations* (pp. 373–383). Springer International Publishing.

[2]Sachan, D. (2018). Self-help robots drive blues away. *Lancet Psychiatry*, *5*(7), 547. https://doi.org/10.1016/s2215-0366(18)30230-x

[3]Landwehr, J. (2023, May 13). People are using ChatGPT in place of therapy—What do mental health experts think? Retrieved 28 September 2023, from https://www.health.com/chatgpt-therapy-mental-health-experts-weigh-in-7488513

conversations with users, helping them recognise their emotions and thought patterns and offering them coping techniques.[4] Researchers suggest that with an adequate approach and sufficient research, MHCBs could be used in psychiatric treatment in the near future.[5] A natural question, then, is whether using MHCBs in a clinical context is a welcome development or rather is ethically worrisome. In this paper, we speak of a 'clinical context' when MHCBs are used in a professional mental health treatment that is aimed at people with diagnosed mental health disorders and where they operate alongside licensed mental healthcare professionals.

On the one hand, MHCBs hold a lot of potential, including improved access to mental health care when there are barriers, such as long waiting lists, shortage of therapists, limited services, or prohibitive costs. In addition, MHCBs are increasingly able to simulate real-life human interaction, which may facilitate not only the quality of treatment but also adherence to treatment compared to other forms of eHealth.[6] MHCBs may also offer benefits compared to regular face-to-face treatment due to the so-called 'online disinhibition effect,' which is the phenomenon that patients tend to self-disclose and express themselves more honestly and openly in the absence of face-to-face contact.[7]

At the same time, however, many authors express ethical concerns, such as concerns over privacy, transparency, responsibility, and accountability, the increase of healthcare inequalities, deception because users tend to anthropomorphise chatbots, and potential harms for both patients (such as safety in case of emergencies) and mental healthcare workers (job loss and marginalisation of therapeutic personnel).[8] This paper addresses a critical yet understudied concern: assuming that MHCBs cannot have genuine emotions, how this assumption may affect psychotherapy, and consequently the quality of treatment outcomes.

In the next section, we provide some background information on how chatbots work. Subsequently, we go into the assumption that MHCBs lack emotions and delve into aspects of the therapeutic process that hinge on therapists' emotions, such as genuine (affective) empathy and countertransference. We articulate some misgivings about the use of MHCBs, given their assumed lack of emotions, especially if patients would otherwise make use of a human therapist. We conclude that further empirical research is necessary to substantiate or assuage these misgivings. Finally, we will go into how our argument contributes to the broader ethical discourse surrounding the widespread use of MHCBs.

2 | HOW DO MENTAL HEALTH CHATBOTS WORK?

A recent study found 18 MHCBs in app stores, operating independently from human therapists. Half of these use artificial intelligence (AI) techniques, such as natural language processing (NLP), and various machine learning methods, to respond to user input using natural language. NLPs are techniques to make computers manipulate and use natural human language, often using machine learning.[9] Machine learning refers to various methods that enable algorithms to learn. The most linguistically human-like chatbots are so-called generative models, which use a type of machine learning called deep learning. In deep learning, algorithms learn directly from raw data without human guidance.[10] ChatGPT and GPT-4 are examples of generative chatbots that use deep learning.[11]

To give an example of how MHCBs work, we will discuss a fairly well-studied MHCB: Woebot, which uses NLP.[12] It aims to reduce depressive and anxiety symptoms using cognitive-behavioural therapy (CBT). As such, it is one of the 14 out of the 18 MHCBs that use CBT techniques.[13] Woebot's conversational style mirrors clinical decision-making, incorporating therapeutic features such as empathic responses.[14] Woebot can be communicated with using a smartphone app. To engage

[4]Fiske, A., Henningsen, P., & Buyx, A. (2019). Your robot therapist will see you now: Ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy. *Journal of Medicine Internet Research*, *21*(5), e13216. https://doi.org/10.2196/13216

[5]Vaidyam, A. N., Wisniewski, H., Halamka, J. D., Kashavan, M. S., & Torous, J. B. (2019). Chatbots and conversational agents in mental health: A review of the psychiatric landscape. *The Canadian Journal of Psychiatry*, *64*(7), 456–464. https://doi.org/10.1177/0706743719828977

[6]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 Ment Health*, *4*(2), e19. https://doi.org/10.2196/mental.7785; Kretzschmar, K., Tyroll, H., Pavarini, G., Manzini, A., & Singh, I. (2019). Can your phone be your therapist? Young people's ethical perspectives on the use of fully automated conversational agents (chatbots) in mental health support. *Biomedical Informatics Insights*, *11*, 1178222619829083. https://doi.org/10.1177/1178222619829083

[7]Berger, T. (2017). The therapeutic alliance in internet interventions: A narrative review and suggestions for future research. *Psychotherapy Research*, *27*(5), 511–524. https://doi.org/10.1080/10503307.2015.1119908; Suler, J. (2004). The online disinhibition effect. *Cyberpsychology Behaviour*, *7*(3), 321–326. https://doi.org/10.1089/1094931041291295

[8]Luxton, D. D. (2014). Recommendations for the ethical use and design of artificial intelligent care providers. *Artificial Intelligence in Medicine*, *62*(1), 1–10. https://doi.org/10.1016/j.artmed.2014.06.004; Luxton, D. D., Anderson, S. L., & Anderson, M. (2016). Ethical issues and artificial intelligence technologies in behavioral and mental health care. In: D. D. Luxton (Eds.), Artificial intelligence in behavioral and mental health care (pp. 255–276). Elsevier. https://doi.org/10.1016/B978-0-12-420248-1.00011-8; Fiske et al., op. cit. note 4, p. e13216; Miller, E., & Polson, D. (2019). Apps, avatars, and robots: The future of mental healthcare. *Issues in Mental Health Nursing*. *40*(3), 208–214. https://doi.org/10.1080/01612840.2018.1524535; Blease, C., Locher, C., Leon-Carlyle, M., & Doraiswamy, M. (2020). Artificial intelligence and the future of psychiatry: Qualitative findings from a global physician survey. *Digital Health*, *6*, 2055207620968355. https://doi.org/10.1177/2055207620968355; Vilaza, G. N., & McCashin, D. (2021). Is the automation of digital mental health ethical? Applying an ethical framework to chatbots for cognitive behaviour therapy. *Frontiers of Digital Health*, *3*, 689736. https://doi.org/10.3389/fdgth.2021.689736; Denecke, K., Abd-Alrazaq, A., & Househ, M. (2021). Artificial intelligence for chatbots in mental health: Opportunities and challenges. Lecture Notes in Bioengineering. Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/978-3-030-67303-1_10; Skorburg, J. A., & Yam, J. (2021). Is there an app for that?: Ethical issues in the digital mental health response to COVID-19. *AJOB Neuroscience*, *13*, 1–14. https://doi.org/10.1080/21507740.2021.1918284; Zhong, M., Bilal, A. M., Papadopoulos, F. C., & Castellano, G. (2021). Psychiatrists' views on robot-assisted diagnostics of peripartum depression. In *Social Robotics: 13th International Conference, ICSR 2021, Singapore, Singapore, November 10–13, 2021, Proceedings* (pp. 464–474). Springer-Verlag. https://doi.org/10.1007/978-3-030-90525-5_40; Sedlakova, J., & Trachsel, M. (2022). Conversational artificial intelligence in psychotherapy: A new therapeutic tool or agent? *The American Journal of Bioethics*, *23*(5), 4–13. https://doi.org/10.1080/15265161.2022.2048739

[9]Adamopoulou and Moussiades, op. cit. note 1, pp. 373–383.

[10]Graham, S., Depp, C., Lee, E. E., Nebeker, C., Tu, X., Kim, H. C., & Jeste, D. V. (2019). Artificial intelligence for mental health and mental illnesses: An overview. *Current Psychiatry Reports*, *21*(11), 116. https://doi.org/10.1007/s11920-019-1094-0

[11]OpenAI. (2023, March 27). *GPT-4 technical report*. arXiv. https://doi.org/10.48550/arXiv.2303.08774

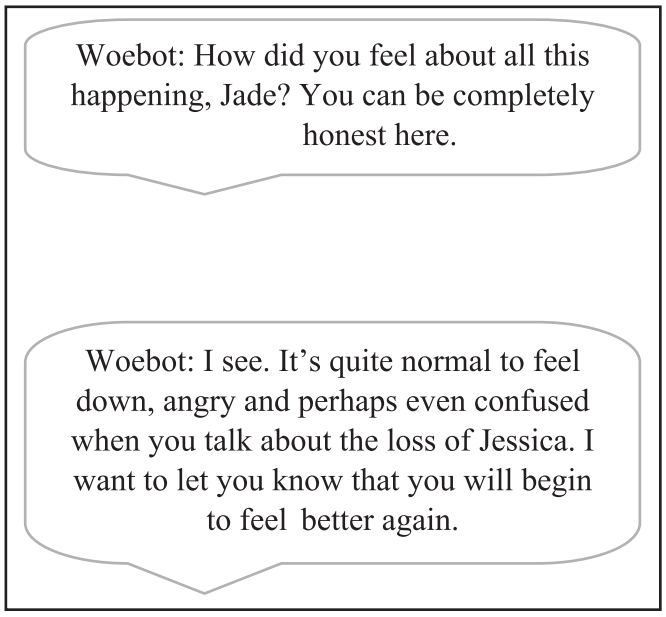
[12]Woebot Health. (n.d.). *What powers Woebot*. Retrieved 11 October 2023, from https://woebothealth.com/what-powers-woebot/

[13]Lin, X., Martinengo, L., Jabir, A. I., Ho, A. H. Y., Car, J., Atun, R., & Tudor Car, L. (2023). Scope, characteristics, behavior change techniques, and quality of conversational agents for mental health and well-being: Systematic assessment of apps. *Journal of Medical Internet Research*. *25*, e45984. https://doi.org/10.2196/45984

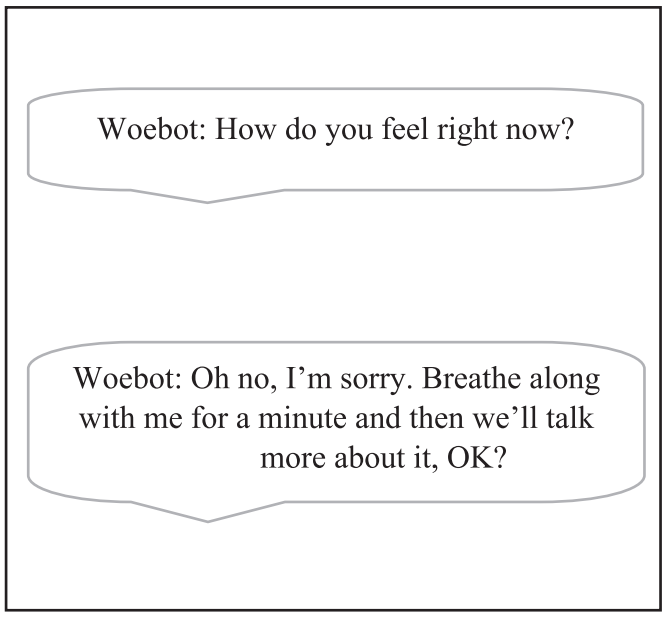
[14]Fitzpatrick et al., op. cit. note 6, p. e19.

in a conversation with Woebot, the user types a text message in a user interface that resembles common chat applications such as WhatsApp. Woebot provides informative answers, supportive comments, and can even explain techniques, such as how users can soothe themselves in cases of anxiety.

Below are two examples of interactions between the user and the MHCB to illustrate that chatbots like Woebot respond in ways that resemble how a human therapist might respond to a patient. The interaction between the user and Woebot proceeds as follows:



Or consider the following exchange:



After this last response, an animation of Woebot shows keeping a regular and calm breathing rhythm.[15]

3 | THE WORRY FROM EMOTIONS

Having understood the structure of a standard MHCB conversation, we will now delve into the central concern of this paper: MHCBs do not possess emotions. If one accepts the assumption that MHCBs lack emotions, this leads to what we will call "The Worry from Emotions"—namely:

(1) If MHCBs lack emotions, they cannot possess genuine (affective) empathy, or utilise countertransference.

[15]Meet Woebot! (2022). https://www.youtube.com/watch?v=ZGBtQw3_Pbo

(2) If MHCBs cannot possess genuine empathy or utilise counter-transference, there are reasons to worry that MHCBs are (a) more liable to harm and (b) less likely to benefit patients than human therapists.
(3) Assumption: MHCBs lack emotions.
(4) So, there are reasons to worry that MHCBs are (a) more liable to harm and (b) less likely to benefit patients than human therapists.

In what follows, we give some reasons to accept (1) and (2) and then articulate how future research should proceed on this topic to more fully determine whether we should accept (1), (2), and (3)—and if so, what that means for the ethics of using MHCBs in a clinical setting.

4 | EMPATHY AND MENTAL HEALTH CHATBOTS

The first aspect of therapy dependent on therapists' emotions is empathy. Empathy is traditionally considered to be an important element of the therapeutic relationship and therefore a cornerstone of psychotherapy.[16] Empathy is amongst the 'common factors' contributing significantly to therapeutic outcome, regardless of therapy type.[17] This includes CBT, which is utilised in most MHCBs.[18] Elliott et al. found in their meta-analysis of 80 studies with a total of 6138 patients (averaging 25 therapy sessions) that a therapist's empathy is a moderately strong predictor of treatment outcomes in psychotherapy. This relation held across different theoretical orientations, with empathy generally accounting for about 9% of therapy outcome variance,[19] which is more than specific treatment methods.[20]

Empathy lacks a consensus definition in psychotherapy,[21] with Cuff et al. finding 43 different definitions.[22] We will not adhere to one specific definition here. However, since most research on empathy in psychotherapy follows Carl Rogers' definition,[23] we will provide it as an example:

> To sense the client's private world as if it were your own, but without ever losing the "as if" quality—this is empathy, and this seems essential to therapy. To sense the client's anger, fear, or confusion as if it were your own, yet without your own anger, fear, or confusion getting bound up in it, is the condition we are endeavouring to describe.[24]

An example would be a therapist noticing sadness developing in themselves while listening to and trying to understand what the patient has gone through at the funeral of a loved one. These feelings can help understand the patient, provided that the therapist's own emotions do not have such a strong impact on them that they hinder helping and understanding the patient. An example of such hindrance would be a therapist that becomes so overwhelmed with sadness from hearing the patient's suffering that they end up crying, unable to discuss the patient's problems productively or even being consoled by the patient for the remainder of the session.

Cuff et al. argue that the 43 different definitions of empathy vary across different themes, such as whether it is congruent or incongruent with the patient's feelings or whether it is cognitive or affective.[25] The relevant question for this paper is whether empathy is cognitive, affective, or both. Cuff et al. note that cognitive empathy is the ability to understand another's feelings, while affective empathy is concerned with the experience of emotion, evoked by an emotional stimulus. They demonstrate that only six out of 43 definitions focus solely on the cognitive component, whereas most are based on the affective component or a combination of both. They conclude that the widely accepted view is that empathy is an emotional event and that in situations where there is no affective element present, we should use another term, such as empathic understanding.[26]

The view that empathy requires the capacity for emotions aligns with neuroscientific findings on empathy's neuroanatomical basis. Empathy involves an *emotional simulation* process which concerns mirroring the emotional components of the other person's bodily experience. It is localised partly in the limbic system, which is involved in emotional processing.[27]

As stated above, empathy is thought to be central to psychotherapy. Patients also prefer therapists with a 'human touch' and a person who can relate to their issues.[28] Considering empathy's affective component, if MHCBs lack emotions, then they lack the capacity for genuine empathy. MHCBs may be designed to provide

[16]Rogers, C. R. (1951). Client-centered therapy: Its current practice, implications, and theory. Houghton Mifflin Co; Norcross, J. C. (2010). The therapeutic relationship. In *The heart and soul of change: Delivering what works in therapy* (2nd ed., pp. 113–141). American Psychological Association. https://doi.org/10.1037/12075-004
[17]Wampold, B. E. (2015). How important are the common factors in psychotherapy? An update. *World Psychiatry*, *14*(3), 270–277. https://doi.org/10.1002/wps.20238
[18]Keijsers, G. P., Schaap, C. P., & Hoogduin, C. A. (2000). The impact of interpersonal patient and therapist behavior on outcome in cognitive-behavior therapy. A review of empirical studies. *Behavior Modification*, *24*(2), 264–297. https://doi.org/10.1177/0145445500242006; Thwaites, R., & Bennett-Levy, J. (2007). Conceptualizing empathy in cognitive behaviour therapy: Making the implicit explicit. *Behavioural and Cognitive Psychotherapy*, *35*(5), 591–612. https://doi.org/10.1017/S1352465807003785
[19]Elliott, R., Bohart, A. C., Watson, J. C., & Murphy, D. (2018). Therapist empathy and client outcome: An updated meta-analysis. *Psychotherapy*, *55*(4), 399–410. https://doi.org/10.1037/pst0000175
[20]Wampold, op. cit. note 17, pp. 270–277.
[21]Elliott et al., op. cit. note 19, pp. 399–410.
[22]Cuff, B. M. P., Brown, S. J., Taylor, L., & Howat, D. J. (2014). Empathy: A review of the concept. *Emotion Review*, *8*, 144–153. https://doi.org/10.1177/1754073914558466
[23]Norcross, op. cit. note 16, pp. 113–141.
[24]Rogers, C. R. (1957). The necessary and sufficient conditions of therapeutic personality change. *Journal of Consulting Psychology*, *21*(2), 95–103. https://doi.org/10.1037/h0045357
[25]Cuff et al., op. cit. note 22.
[26]Ibid.
[27]Elliott et al., op. cit. note 19, pp. 399–410.
[28]Salamanca-Sanabria, A., Jabir, A. I., Lin, X., Alattas, A., Kocaballi, A. B., Lee, J., Kowatsch T., & Car, L. T. (2023). Exploring the perceptions of mHealth interventions for the prevention of common mental disorders in university students in Singapore: Qualitative study. *Journal of Medical Internet Research*, *25*(1), e44542. https://doi.org/10.2196/44542

responses perceived as empathic by users,[29] and a recent study indeed found that ChatGPT was perceived to provide more empathic responses to health questions on an online forum than verified human physicians.[30] Nonetheless, these responses are simulations of empathic responses, rather than genuine empathy. This may raise the worry that, when it comes to empathy, a lesser quality of psychotherapeutic care can be expected from MHCBs because they cannot benefit from the positive outcomes of genuine affective empathy.

5 | COUNTERTRANSFERENCE AND MENTAL HEALTH CHATBOTS

Yet, one might argue that MHCBs can still provide effective care even without emotions. They can offer feedback or advice similar to human therapists, such as helping users identify and restructure cognitive distortions. For example, Woebot's modelling of appropriate breathing rates could be as helpful as a human therapist's intervention, even without emotions. Furthermore, some may contend that as long as MHCBs can simulate empathy in therapy, patients will benefit. Perhaps the perception of a 'human touch' suffices, and some patients may perceive MHCBs as possessing this quality.

While this reply to concerns related to the absence of emotions in MHCBs may have some merit, one complication is that emotions are necessary for *countertransference*.[31] Countertransference builds on the concept of transference. Transference refers to the strong feelings that the patient experiences, which are not created in the therapeutic process but rather are feelings towards past role models that are transferred onto the therapist.[32] The role of transference has already been addressed as potentially problematic in the use of MHCBs as therapists.[33] Today, countertransference encompasses both conscious and unconscious, and both internal and external reactions of the therapist to the patient, based not only on the patient's transference but also on the therapist's own unresolved conflicts.[34]

An important distinction between countertransference and therapist empathy is that empathetic feelings typically align with those of the other person, described as 'feeling as if'.[35] Countertransference, however, does not require such congruence.[36] A significant proportion of the variance in the therapist's countertransference responses is attributable to the patient.[37] This suggests a purported therapeutic benefit: human therapists gain insights into their patients through examining their own countertransference responses.[38]

Colli and Ferri's 2015 review offers an overview of empirical studies on the relationship between therapist countertransference reactions and patient diagnoses. They conclude, while acknowledging limitations, that multiple studies demonstrate a relationship between specific therapist reactions and patient personality disorder types or clusters. And that studies suggest that these therapist responses occur across any kind of therapy, regardless of the therapist's theoretical orientation.[39] Stefana et al.'s 2020 systematic review reached similar conclusions, suggesting therapist reactions towards patients provide valuable diagnostic information.[40]

Empirical studies and a systematic review across various psychotherapeutic approaches, such as short- and long-term psychodynamic psychotherapy, cognitive behavioural group therapy, psychoanalytic psychotherapy and psychological counselling, all indicate that positive countertransference (which is the representation of positive feelings towards the patient such as closeness, interest, and respect) is associated with positive outcomes such as symptom improvement.[41] In sum, countertransference is considered important for psychotherapy,[42] including CBT.[43]

[29]Inkster, B., Sarda, S., & Subramanian, V. (2018). An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: Real-world data evaluation mixed-methods study. *JMIR Mhealth Uhealth*, *6*(11), e12106. https://doi.org/10.2196/12106; Morris, R. R., Kouddous, K., Kshirsagar, R., & Schueller, S. M. (2018). Towards an artificially empathic conversational agent for mental health applications: system design and user perceptions. *Journal of Medical Internet Research*, *20*(6), e10148. https://doi.org/10.2196/10148

[30]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

[31]Racker, H. (2018). *Transference and countertransference*. Routledge.

[32]Gabbard, G. O. (2014). *Psychodynamic psychiatry in clinical practice* (5th ed.). American Psychiatric Publishing.

[33]Holohan, M., & Fiske, A. (2021). 'Like I'm talking to a real person': Exploring the meaning of transference for the use and design of ai-based applications in psychotherapy. *Frontiers of Psychology*, *12*, 720476. https://doi.org/10.3389/fpsyg.2021.720476

[34]Hayes, J. A., Gelso, C. J., Goldberg, S., & Kivlighan, D. M. (2018). Countertransference management and effective psychotherapy: Meta-analytic findings. *Psychotherapy*, *55*(4), 496–507. https://doi.org/10.1037/pst0000189; Hayes, J. A., Nelson, D. L. B., & Fauth, J. (2015). Countertransference in successful and unsuccessful cases of psychotherapy. *Psychotherapy*, *52*(1), 127–133. https://doi.org/10.1037/a0038827

[35]Cuff et al., op. cit. note 22.

[36]Rossberg, J. I., Karterud, S., Pedersen, G., & Friis, S. (2010). Psychiatric symptoms and countertransference feelings: An empirical investigation. *Psychiatry Research*, *178*(1), 191–195. https://doi.org/10.1016/j.psychres.2009.09.019; Sandler, J., Holder, A., & Dare, C. (1970). Basic psychoanalytic concepts. IV. Counter-transference. *The British Journal of Psychiatry: The Journal of Mental Science*, *117*(536), 83–88.

[37]Löffler-Stastka, H., Sell, C., Zimmermann, J., Huber, D., & Klug, G. (2019). Is countertransference a valid source of clinical information? Investigating emotional responses to audiotaped psychotherapy sessions. *Bulletin of the Menninger Clinic*, *83*(4), 353–375. https://doi.org/10.1521/bumc_2019_83_02

[38]Gabbard, G. O. (2020). The role of countertransference in contemporary psychiatric treatment. *World Psychiatry*, *19*(2), 243–244. https://doi.org/10.1002/wps.20746; Colli, A., & Ferri, M. (2015). Patient personality and therapist countertransference. *Current Opinion in Psychiatry*, *28*(1), 46–56. https://doi.org/10.1097/YCO.0000000000000119

[39]Colli & Ferri, op. cit. note 38, pp. 46–56.

[40]Stefana, A., Bulgari, V., Youngstrom, E. A., Dakanalis, A., Bordin, C., & Hopwood, C.J. (2020). Patient personality and psychotherapist reactions in individual psychotherapy setting: A systematic review. *Clinical Psychology & Psychotherapy*, *27*(5), 697–713. https://doi.org/10.1002/cpp.2455

[41]Rossberg et al., op. cit. note 36, pp. 191–195; Machado, D. de B., Coelho, F. M. da C., Giacomelli, A. D., Donassolo, M. A. L., Abitante, M. S., Dall'Agnol, T., & Eizirik, C. L. (2014). Systematic review of studies about countertransference in adult psychotherapy. *Trends in Psychiatry and Psychotherapy*, *36*, 173–185. https://doi.org/10.1590/2237-6089-2014-1004; Gazzillo, F., Lingiardi, V., Del Corno, F., Genova, F., Bornstein, R. F., Gordon, R. M., & McWilliams, N. (2015). Clinicians' emotional responses and Psychodynamic Diagnostic Manual adult personality disorders: A clinically relevant empirical investigation. *Psychotherapy*, *52*(2), 238–246. https://doi.org/10.1037/a0038799; Löffler-Stastka et al., op. cit. note 37, pp. 353–375.

[42]Gabbard, op. cit. note 38, pp. 243–244.

[43]Gluhoski, V. L. (1994). Misconceptions of cognitive therapy. *Psychotherapy: Theory, Research, Practice, Training*, *31*(4), 594–600. https://doi.org/10.1037/0033-3204.31.4.594; Prasko, J., Ociskova, M., Vanek, J., Burkauskas, J., Slepecky, M., Bite, I., Krone, I., Sollar, T., & Juskiene, A. (2022). Managing transference and countertransference in cognitive behavioral supervision: Theoretical framework and clinical application. *Psychology Research and Behavior Management*, *15*, 2129–2155. https://doi.org/10.2147/PRBM.S369294

If we assume that MHCBs lack emotions, even if they simulate them,[44] then they are incapable of experiencing countertransference responses. This gives reason to worry that patients using MHCBs may not benefit from the potential positive outcomes of countertransference responses. This can be particularly challenging in certain patient groups and therapy types. For example, research suggests that certain personality disorders are more specifically associated with certain countertransference responses.[45]

Thus, if MHCBs lack genuine empathy and countertransference and therefore cannot benefit from their potential positive outcomes, this is a reason to think they will be (a) more liable to harm and (b) less likely to benefit patients than human therapists. However, to assess the validity of these concerns, we must consider additional arguments, which will be addressed in the following section.

6 | MENTAL HEALTH CHATBOTS: MORE LIABLE TO HARM AND LESS LIKELY TO BENEFIT PATIENTS?

To recapitulate, we have given some reasons to accept the first steps in The Worry from Emotions—(1): if MHCBs lack emotions, they cannot possess genuine (affective) empathy or utilise countertransference. We have consequently contended that (2) if MHCBs cannot possess genuine empathy or utilise countertransference, there are reasons to worry that MHCBS are (a) more liable to harm and (b) less likely to benefit patients than human therapists. However, there are some counterarguments that need to be taken into account.

First, the lack of emotions in MHCBs may yield certain advantages compared to human therapists. As mentioned earlier, the online disinhibition effect leads some patients to be more honest and open in online environments than when meeting face-to-face with a human therapist. The absence of emotions in MHCBs may offer similar advantages relative to human therapists. Patients may feel less judged by MHCBs, potentially leading to greater honesty in disclosing their problems, or patients might be less likely to attempt to please or entertain their human therapist. For instance, a study among military servicemen showed that they disclosed more mental health symptoms to a chatbot than on an anonymous self-report checklist, where anonymity is expected to encourage disclosure.[46] To determine the degree to which the absence of emotions constitutes a net impediment to good care, more research is needed into whether the lack of emotions on the part of MHCBs can also yield advantages in providing good care and how this counterbalances the disadvantages of this lack.

Second, the fact that human therapists have emotions whereas MHCBs lack them may make *human* therapists more liable to harm and less likely to benefit patients. For example, MHCBs will not pursue inappropriate relationships with patients—they cannot, after all, form inappropriate emotional attachments to their patients or even have sex with their patients, whereas human therapists are all-too-capable of such misconduct.[47]

Indeed, countertransference is not only regarded as a valuable source of information about the patient but also sometimes as an obstruction to therapy.[48] For example, because of so-called 'negative' countertransference and their own emotional history, human therapists may sometimes inappropriately respond to patients—thereby causing harm. Norris et al. argue that therapists could disclose that they are sexually aroused to a patient as a misjudged countertransference-based intervention.[49] This shows how certain countertransference feelings (such as being sexually aroused) could be potentially damaging to a patient (such as crossing boundaries and creating an unsafe therapeutic environment), when they are not adequately reflected on and managed by the therapist. Studies have shown that better countertransference management is associated with better psychotherapy outcome.[50] Thus, having the requisite capacities for emotions also entails some risk of harming patients.

To recapitulate, it seems plausible enough that the emotional limitations of MHCBs might well make them worse than an ideal human therapist. However, not all human therapists meet this ideal. To fully determine the impact of MHCBs' lack of emotions on quality of care, more research is needed into the extent to which the *possession* of emotions in human therapists can also yield *disadvantages* to providing good care as compared to MHCBs.

So far, we discussed concerns stemming from MHCBs' lack of emotions regarding their ability to have empathy and to utilise countertransference. We have identified some areas that need further empirical research to find out whether these concerns are valid. In the following section, we will reflect on how our considerations contribute to the broader ethical discourse surrounding the widespread use of MHCBs.

7 | THE BROADER ETHICAL DEBATE ON THE CLINICAL USE OF MENTAL HEALTH CHATBOTS

While our analysis highlights critical areas that necessitate further investigation with MHCBs, it is essential to contextualise these within the broader landscape of ethical considerations surrounding their widespread implementation in a clinical context.

[44]Weber-Guskar, E. (2021). How to feel about emotionalized artificial intelligence? When robot pets, holograms, and chatbots become affective partners. *Ethics and Information Technology*, *23*(4), 601–610. https://doi.org/10.1007/s10676-021-09598-8

[45]Gazzillo et al., op. cit. note 41, pp. 238–246.

[46]Lucas, G. M., Rizzo, A., Gratch, J., Scherer, S., Stratou, G., Boberg, J., & Morency, L.-P. (2017). Reporting mental health symptoms: Breaking down barriers to care with virtual human interviewers. *Frontiers in Robotics and AI*, *4*, 0. https://www.frontiersin.org/articles/10.3389/frobt.2017.00051

[47]Norris, D. M., Gutheil, T. G., & Strasburger, L. H. (2003). This couldn't happen to me: Boundary problems and sexual misconduct in the psychotherapy relationship. *Psychiatric Services*, *54*(4), 517–522. https://doi.org/10.1176/appi.ps.54.4.517

[48]Parth, K., Datz, F., Seidman, C., & Löffler-Stastka, H. (2017). Transference and countertransference: A review. *Bulletin of the Menninger Clinic*, *81*(2), 167–211. https://doi.org/10.1521/bumc.2017.81.2.167

[49]Ibid.

[50]Hayes et al., op. cit. note 34, pp. 496–507.

Despite requiring more research to know whether our concerns are valid, let us provisionally grant that because of their lack of emotions, MHCBs are (a) more liable to harm and (b) less likely to benefit patients than human therapists, that is, let us grant The Worry from Emotions. This leads to a final concern: if MHCBs are more liable than human therapists to (a) and (b) Is their widespread clinical adoption ethically worrisome? To answer this, we should not only address concerns about emotions but also consider broader ethical debates on the use of MHCBs, including the concerns raised earlier in the introduction.

For one thing, we might still argue that MHCBs provide a net benefit to many individuals—despite their lack of emotions. One way they might provide a net benefit is if, as others have proposed,[51] patients use MHCBs *alongside* human therapists. For instance, perhaps the MHCB can assist with exercises at home where the patient does not have access to a human therapist. Additionally, MHCBs could be used for tasks that do not rely heavily on empathy and countertransference, such as psychoeducation, thereby freeing up time for the therapist. Psychoeducation, a fundamental element of overall psychological treatment, involves explaining symptoms, causes, and treatment options to patients and their relatives. The vast amount of information that MCHBs have access to could be very useful for psychoeducation.

Furthermore, while the average patient would benefit more from a human therapist than using a MHCB, there could be individuals who fare better with a MHCB, for reasons discussed earlier. Such patients might be less inhibited, less inclined to lie to their therapist, or less motivated to try to please their therapist, hence they might benefit from the online disinhibition effect. Empirical studies are necessary to reveal whether the online disinhibition effect is indeed beneficial with MHCBs and whether these effects counterbalance the potential disadvantages of MHCBs not having emotions in certain patient groups. Related to this topic, further research is required to explore the consequences of users anthropomorphising chatbots and the potential for dependency on them.[52]

A final question concerns just how substandard the care from a MHCB must be before we evaluate it as ethically troublesome. This question pertains to the baseline for sufficiently good care. One way to articulate this worry is through the on-going debate about the point at which care that falls short of the 'gold standard' may still be provided—especially in contexts where the unavailability or high cost of optimal treatment motivates the use of suboptimal treatment.[53] To illustrate, Persad and Emanuel argue against the WHO that it is ethically appropriate to provide less effective or more dangerous anti-retroviral therapies in contexts where the gold standard of ART is not available and that it is permissible to donate medical technologies known to be less effective than the gold standard.[54] We certainly do not intend to weigh in on this debate here. Rather, we wish to note that there is a parallel question: if we find that the quality of care provided by MHCBs is lower than the care provided by human therapists, but the 'gold standard of care' (i.e., a human therapist) would not otherwise be available at all, does it follow that it is ethically worrisome if individuals use MHCBs instead?

8 | CONCLUSION

This paper has explored ethical concerns related to the limitations of a rapidly developing piece of technology: MHCBs. These chatbots are already being commercially used by individuals facing mental health problems. There is a growing body of literature addressing various ethical issues regarding the use of MHCBs. This paper has examined the desirability of using MHCBs in clinical settings and delves into a less-explored area of concern. Building on the assumption that MHCBs lack emotions, this paper has explored ethical concerns arising from this limitation, and whether it renders them unfit to provide psychotherapy.

We have articulated some ways in which MHCBs cannot replicate qualities of human psychotherapists that relate to having emotions. First, we looked at empathy, an important element of psychotherapy and a contributor to treatment outcomes. We have argued that if MHCBs lack emotions, they cannot possess genuine (affective) empathy. Second, we have examined countertransference, an element of psychotherapy that therapists use to gain insights into the patients and that also contributes treatment outcomes. We have consequently argued that if MHCBs lack emotions, they are unable to utilise countertransference. This lack of genuine empathy and countertransference raises concerns about whether MHCBs are (a) more liable to harm and (b) less likely to benefit patients than human therapists.

Further research is needed to fully assess whether these concerns are valid. Future research avenues include exploring whether the lack of emotions in MHCBs could also be beneficial for certain patient groups and whether the possession of emotions can constitute a disadvantage for human therapists relative to MHCBs. Finally, we have touched on the broader ethical debate surrounding the use of MHCBs, exploring questions about whether their use alongside human therapists can be justified and when substandard care can be warranted. Addressing these questions will be instrumental in shaping the responsible and effective deployment of MHCBs while ensuring the well-being and safety of patients in need of mental health support.

[51]Luxton, op. cit. note 8, pp. 1–10; Luxton et al., op. cit. note 8, pp. 255–276; Fiske et al., op. cit. note 4, p. e13216; Miller & Polson, op. cit. note 8, pp. 208–214; Blease et al., op. cit. note 8, p. 2055207620968355; Fiske, A., Henningsen, P., & Buyx, A. (2020). The implications of embodied artificial intelligence in mental healthcare for digital wellbeing. *Philosophical studies series* (Vol. 140). Springer Nature. https://doi.org/10.1007/978-3-030-50585-1_10; Brown, C., Story, G. W., Mourão-Miranda, J., & Baker, J. T. (2021). Will artificial intelligence eventually replace psychiatrists? *The British Journal of Psychiatry*, *218*(3), 131–134. https://doi.org/10.1192/bjp.2019.245; Vilaza & McCashin, op. cit. note 8, p. 689736; Omarov, B., Narynov, S., & Zhumanov, Z. (2022). Artificial intelligence-enabled chatbots in mental health: A systematic review. *Computers, Materials & Continua*, *74*(3), 5105–5122. https://doi.org/10.32604/cmc.2023.034655; Sedlakova &Trachsel, op. cit. note 8, pp. 1–10; Al-Ameery-Brosche, I., & Resch, F. (2022). Emotional robotics: Curse or blessing in psychiatric care? In R. M. Holm-Hadulla, J. Funke, & M. Wink (Eds.), *Intelligence—Theories and applications* (pp. 261–271). Springer International Publishing. https://doi.org/10.1007/978-3-031-04198-3_15

[52]Whitby, B. (2014). The ethical implications of non-human agency in health care. In *AISB 2014—50th Annual Convention of the AISB*. Society for the Study of Artificial Intelligence and the Simulation of Behaviour; Sedlakova & Trachsel, op. cit. note 8, pp. 1–10.

[53]Persad, G. C., & Emanuel, E. J. (2017). The case for resource sensitivity: Why it is ethical to provide cheaper, less effective treatments in global health. *Hastings Center Report*, *47*(5), 17–24. https://doi.org/10.1002/hast.764

[54]Ibid: 17–18.

ACKNOWLEDGEMENTS

An earlier version of this paper was presented at the Machines of Change: Robots, AI, and Value Change Workshop of the TU Delft and has benefited from the discussions there. We would also like to thank colleague and friend Devin Sanchez Curry for his valuable feedback on our manuscript.

ORCID

Mehrdad Rahsepar Meadi http://orcid.org/0000-0001-5637-3349
Suzanne Metselaar http://orcid.org/0000-0001-9214-6477

AUTHOR BIOGRAPHIES

Mehrdad Rahsepar Meadi is a PhD candidate affiliated with the Department of Psychiatry and the Department of Ethics, Law & Humanities of Amsterdam UMC, location VU University. He is also a psychiatry resident at the mental health organisation GGZ in Geest in Amsterdam. In addition to his medical training, he has completed a master's degree in bioethics. His PhD project concerns the ethics of AI-driven conversational agents in mental health care.

Justin S. Bernstein, PhD, is an assistant professor of Ethics and Political Philosophy at Vrije Universiteit, Amsterdam. His research focuses on various topics in public health ethics and political philosophy.

Neeltje Batelaan is professor of anxiety-related disorders at Amsterdam UMC, Amsterdam, the Netherlands, and chair of the Dutch Knowledge Center for anxiety, OCD, and depression (NEDKAD). As a psychiatrist, she is working in an outpatient department for anxiety disorders.

Anton J. L. M. van Balkom, MD, PhD (1956) specialises in evidence-based psychiatry. He has conducted meta-analyses, randomised controlled trials and implementation studies in obsessive compulsive disorders, and anxiety disorders which has led to more than 200 international scientific publications and over 25 PhD dissertations. He is chair of the Netherlands multidisciplinary guidelines committee for anxiety and obsessive-compulsive disorders.

Dr. Suzanne Metselaar is a medical ethicist and senior researcher with a PhD in philosophy. Her research focuses on quality of care, the philosophical foundations of clinical ethics, and palliative care.

How to cite this article: Rahsepar Meadi, M., Bernstein, J. S., Batelaan, N., van Balkom, A. J. L. M., & Metselaar, S. (2024). Does a lack of emotions make chatbots unfit to be psychotherapists? *Bioethics*, *38*, 503–510. https://doi.org/10.1111/bioe.13299