

# Considering AI-driven therapy: When does human empathy matter?

Matan Rubin, Hadar Arnon, Jonathan Huppert, Anat Perry

Submitted to: JMIR Mental Health  
on: January 18, 2024

**Disclaimer:** © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

*Table of Contents*

---

Original Manuscript..... 4

Supplementary Files..... 22

0..... 23

Preprint  
JMIR Publications

# Considering AI-driven therapy: When does human empathy matter?

Matan Rubin<sup>1</sup>; Hadar Arnon<sup>1</sup>; Jonathan Huppert<sup>1</sup> PhD; Anat Perry<sup>1</sup> PhD

<sup>1</sup>Psychology Department Hebrew University of Jerusalem Jerusalem IL

## Corresponding Author:

Anat Perry PhD

Psychology Department

Hebrew University of Jerusalem

Mt Scopus

Jerusalem

IL

## Abstract

Recent breakthroughs in artificial intelligence (AI) language models have elevated the vision of using AI support for mental health, with a growing body of literature indicating varying degrees of efficacy. In this paper we ask when in therapy will it be easier to replace humans, and conversely, in what instances will human connection still be more valued. We suggest that empathy lies at the heart of the answer to this question. First, we define different aspects of empathy and outline potential empathic capabilities of humans versus AI. Next, we consider what determines when these aspects are needed most in therapy, both from the perspective of therapeutic approach, and from the perspective of patient objectives. Ultimately, our goal is to prompt further investigation and dialogue, urging both practitioners and scholars engaged in AI-mediated therapy to keep these questions and considerations in mind when investigating AI implementation in mental health.

(JMIR Preprints 18/01/2024:56529)

DOI: <https://doi.org/10.2196/preprints.56529>

## Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <http://www.jmir.org>

## Original Manuscript

## **Considering AI-driven therapy: When does human empathy matter?**

Matan Rubin, Hadar Arnon, Jonathan D. Huppert & Anat Perry\*

Psychology Department, The Hebrew University of Jerusalem

\*Corresponding author:

Anat Perry

Psychology Department

Hebrew University of Jerusalem 91905

anat.perry@mail.huji.ac.il

## Abstract

Recent breakthroughs in artificial intelligence (AI) language models have elevated the vision of using conversational AI support for mental health, with a growing body of literature indicating varying degrees of efficacy. In this paper we ask when in therapy will it be easier to replace humans, and conversely, in what instances will human connection still be more valued. We suggest that empathy lies at the heart of the answer to this question. First, we define different aspects of empathy and outline potential empathic capabilities of humans versus AI. Next, we consider what determines when these aspects are needed most in therapy, both from the perspective of therapeutic methodology, and from the perspective of patient objectives. Ultimately, our goal is to prompt further investigation and dialogue, urging both practitioners and scholars engaged in AI-mediated therapy to keep these questions and considerations in mind when investigating AI implementation in mental health.

**Keywords: empathy, artificial empathy, AI, mental health**

## Introduction

The prospect of employing machine learning algorithms for automated healthcare responses and counseling has long been considered [1]. Such algorithms could have vast benefits ranging from increased accessibility and affordability of mental health services, reduced waiting times, and personalized treatment options, to the potential to reach underserved populations and combat the escalating loneliness epidemic [2]. Recent breakthroughs in artificial intelligence (AI) language models have elevated this vision, as evidenced by a growing body of literature indicating varying degrees of efficacy. For instance, studies demonstrate that digital chatbots are proficient in delivering psychoeducation and in improving treatment adherence over short durations [3]. Additionally, AI-driven chatbots have been effectively utilized to impart strategies derived from positive psychology and cognitive-behavioral techniques to mitigate stress and enhance subjective well-being [4]. AI chatbots can also provide preliminary support during the absence of a therapist by prompting self-reflective questioning and facilitating emotion regulation in challenging scenarios [5,6]. Machine learning can also be used in order to detect symptom changes in new ways [7]. In the realm of medicine, a recent study revealed that the responses from GPT-3 received higher ratings for the quality of medical advice compared to physicians. Moreover, these responses were perceived to exhibit significantly more empathy compared to those from physicians [8].

While the potential for AI chatbots to take over certain elements of the therapeutic process exists, there are compelling reasons to believe that they cannot completely substitute for the human element. This raises a critical

question: under what circumstances will human therapists remain indispensable, and conversely, when could they feasibly be replaced by AI models? We suggest that part of the answer may reside in an exploration of the role of empathy in the therapeutic process. In the following paper, we address the multifaceted nature of empathy, including its cognitive, emotional, and motivational aspects. We claim that in those cases where emotional or motivational empathy is needed, humans will be harder to replace. We then consider what determines when these aspects are needed most in therapy—whether certain therapeutic approaches, particular patient goals, or specific points within the therapeutic timeline. Our inquiry explores these considerations from both the perspective of therapeutic methods and patient objectives. Ultimately, our goal is to prompt further investigation and dialogue, urging both practitioners and scholars engaged in AI-mediated therapy to consider these issues through the lens of empathy and human connection.

## Empathy

A comprehensive definition of empathy recognizes three dimensions of empathic engagement: cognitive empathy, or mentalizing, which pertains to the recognition and understanding of the emotional states of others; emotional empathy, or affective sharing, which involves resonating with others' emotional experiences while maintaining self-other differentiation; and motivational empathy, often termed empathic concern or compassion, which encompasses feelings of concern for another's welfare and a readiness to act to enhance their well-being [9].

Current advances in natural language processing and facial recognition

technologies have positioned AI-based algorithms at the forefront of discerning emotional states [10,11], with projections suggesting that they may reach or surpass human capability in the near future. Therefore, in the most basic sense of cognitive empathy as recognition of the other's emotional state, AI algorithms will probably do quite well.

Nonetheless, AI, at least in its current form, does not exhibit the latter two empathic capacities. AI does not partake in emotional experiences—it neither shares in joy nor sorrow. Therefore, regardless of how eloquently it crafts a response to seem like it shares an emotional experience, this response will be untruthful, as it does not share any experience. While such responses may still have some benefits, they will probably not be experienced, by the listener, in the same way [See for example 12,13].

Moreover, conversational AI does not have the capacity to manifest genuine care and concern. Human expressions of empathic care signal a willingness to bear an emotional burden and expend limited cognitive-emotional resources on the interaction. Empathy, being potentially taxing, is selectively directed, often preferentially towards close relations and ingroup members, rather than those more distant [14]. In this way, such expressions signify the recipient's importance and closeness to the empathizer. Indeed, studies show that stripped from context and motivation, individuals often tend to avoid empathy [15,16]. Thus, whether in therapy or in social or professional realms, authentic expressions of empathy are significant to the recipient because they reflect a conscious commitment of time, thought, and emotional labor from the empathizer. Though these resources are inherently scarce for humans, they are unlimited to a conversational AI model. Its response is essentially cost-free, and

it would react with comparable enthusiasm to anyone else. As a result, the conversational AI's empathy fails to convey authentic care or indicate that the recipient holds any unique importance [17].

Empathy, and specifically its emotional and motivational components, has been consistently linked to positive outcomes in treatment. The extent of empathy expressed by the therapist and perceived by the patient has a substantial correlation with the success of the treatment [18]. Rogers [19, p142] even describes the therapeutic process as a mutual participation in an emotional exchange, which is then accurately interpreted and reflected upon with the patient to facilitate understanding of their experiences. As Rogers articulated, comprehending the patient's emotions (cognitive empathy) is imperative for endorsing and designing goals and interventions that confront these emotions. This process is underpinned by a commitment to assist and support the patient (motivational empathy), both of which stem from participating in the patient's emotional journey (affective empathy).

Upon considering the importance of empathy for the therapeutic process and outcomes, as well as the limitations of AI discussed above, several questions arise. First, given the limitations, for which aspects of therapy could AI completely replace human therapists? Second, in aspects of therapy where human empathy is essential, how can AI algorithms assist therapists? For example, could it aid therapists in being more accurate or in being more committed to their patients (perhaps via enhancing therapist understanding and empathy and by potentially reducing burnout)?

While we do not claim to give clear-cut answers, this paper explores these questions through dual lenses: the perspective of the therapeutic approach and

the perspective that prioritizes specific motivations and needs of the patient.

## Perspectives in Psychotherapy

Psychotherapy encompasses a diverse spectrum of approaches. A major debate in the field of psychotherapy concerns the mechanism of change. To state it simplistically, one extreme viewpoint contends that the therapeutic relationship is the main mechanism [e.g., 20], whereas the other extreme argues that therapeutic techniques or procedures are the main mechanism [e.g., 21]. At times, these two stances are reflected in therapeutic approaches such that psychodynamic (i.e., Neo-Freudian) approaches tend to emphasize the therapeutic relationship whereas cognitive-behavioral approaches tend to emphasize technique. In reality, most psychotherapies attempt to integrate some combination of the two and allow them to build on one another: there are techniques used to form the relationship, and relationships facilitate use of techniques. Furthermore, a given act can be seen as both relationship-building and as the administration of a technique. Essentially, the therapeutic process often demands that the therapist engage in a comparative analysis of emotional experiences with the patient, thus exercising some form of affective empathy, though it is possible that these affective components of empathy are more crucial in some therapeutic interventions than in others.

Despite the variance in therapeutic orientations, there is broad agreement that one of the fundamental transtheoretical elements critical for a successful outcome is the treatment's working alliance [22]. Conventionally, the alliance is measured across three domains: agreement on therapeutic goals, agreement on

the therapeutic activities needed to achieve these goals, and the warmth and genuineness of the connection. The working alliance is a significant predictor of treatment outcomes across all forms of psychotherapy [23,24]. Motivational empathy is intimately connected to a fruitful working alliance, particularly the aspect concerning warmth and care [25]. Such results were reported in early studies of cognitive-behavioral therapy, where warmth was a predictor of symptom improvement [26], though subsequent findings have been equivocal [21].

One of the provocative findings regarding the alliance of client and therapist is that it appears to be similarly related to reported outcomes in face-to-face psychotherapy and in online internet-based interventions (IBI) which include asynchronous communications [24] with minimal therapist contact. Within the field of IBI, the presence of therapist support is predictive of more symptom improvement, less dropout, and greater adherence in comparison with unguided IBI [27,28] (though data are not conclusive; see [29]). Furthermore, the relationship with the internet-based program has been found to be predictive of symptom reduction whereas the relationship with the therapist was predictive of adherence and dropout in a therapist-guided IBI [30]. These differences raise the possibility that although patients can potentially form a relationship with a digital interface, such a relationship differs in the benefits in comparison to the relationship with a therapist, and may not be able to create the same profound alliance and conversations with patients [31]. In other words, conversational AI and digital programs can be helpful in psychoeducation and administering practical techniques that alleviate symptoms, but they are unlikely to build a motivational relationship in the same way as a human

therapist and client. Moreover, it has been claimed that even if conversational AI were to be used only for practical purposes alongside a human therapist, this may change and interact with the human dynamic in therapy, especially with the therapeutic alliance [32], which may possibly change how one experiences empathy throughout the therapeutic process.

## **A Patient-Needs Perspective**

The previous section examined the importance of empathy in therapy, from the perspective of therapeutic treatment approaches, which could be thought of as a continuum in terms of their theoretical mechanisms from skill acquisition (e.g., cognitive-behavioral therapy) to relationship-based (e.g., interpersonal psychoanalysis). An alternate perspective considers a patient's specific motivations and needs for seeking treatment, regardless of the therapist's theoretical orientation. We contend that patients enter therapy with a variety of needs, motivations, and expectations, which can be conceptualized along two, orthogonal continua including 1) a desire for acquiring practical tools and 2) a desire for human connection and empathy. The relative emphasis on each dimension varies among patients according to their theory of change (i.e., what they need to decrease suffering or improve quality of life) and many other factors (such as history of successful/unsuccessful treatment, stigma, culture, etc.). In addition to individual variability, we contend that individuals can change in their theory and emphasis during therapy as a result of their cumulative experiences over the course of therapy. The more a patient seeks to acquire strategies to cope, the more conversational AI might be able to facilitate this process by providing psychoeducation, exercises, and the like. Conversely, the

greater the patient's need for human connection and empathy—be it for affirmation, a confidant for their thoughts and feelings, or simply the sense that someone cares—the less capable conversational AI might be in fulfilling these requirements. We note that although we see these as two dimensions, we would not expect many individuals who seek treatment to be low on both, as there would be no motivation to seek treatment (aside from appeasing others).

Adding to this complexity is the patient's self-awareness and accuracy regarding their needs, which is often ambiguous (e.g., they may believe they need practical tools when, in reality, they require empathic care more, or vice versa). It is conceivable that a patient might need coping skills for personal growth, yet the most effective means of acquiring these tools could be through engagement with a compassionate therapist. Such a process may take time and require working on the therapeutic relationship, and needs may change during this period. We believe that when thinking of how to utilize AI systems in mental health contexts we must not seek to use it as a “quick fix” solution to a specific problem at all times, as in some cases a long-term human connection may be required to give more thorough help.

Moreover, even in therapeutic interactions that are not complete courses of psychotherapy and are comprised of short, concrete interventions, such as crisis helplines, people still may need a human connection. Indeed, research shows that the main reason people call crisis helplines is to have someone to talk to [33]. While conversational AI systems can, at times, give the feeling of “being heard”, its responses have still been reported to have less value than those written by a human [13]. One could assume that repeat callers [34] to such helplines seek human connection and will not find it sufficient to only

communicate with a bot. A bot-only approach may lead to deterioration in their condition, a risk that should be minimized, especially if a person is going through a crisis. Additionally, other users develop an overdependency on the AI tool, and interventions should be planned in a manner that would mitigate the risk of a long-term dependency. It is possible that AI-assisted communication including a therapist could help deal with both of these risks.

We bring this population as an example, but these considerations pose open-ended questions that warrant exploration in the burgeoning domain of AI-mediated therapy broadly, where the interplay between human touch and technological aid is continually redefined.



Figure 1: An illustration of different patients' possible needs, and their possible benefit from AI intervention or a human connection in therapy

## Patient Perceptions of AI-Bots

In the growing literature examining individuals' perceptions of AI bots, different

factors have been shown to influence the levels of trust and bonding created with AI [35–38]. Much of this literature can be viewed from the same two axes: how helpful and capable is the bot in providing appropriate tools and results, and how empathic it was (in terms of understanding the user's needs). One study showed that a therapeutic bot was rated on the dimensions of alliance at similar levels to face-to-face therapists [39]. In another study, responses generated by ChatGPT were rated as more authentic, professional, and practical [40], however, participants were blind to the fact that responses were generated by conversational AI. When participants are aware of AI systems' involvement, or simply believe it is involved, responses can seem less authentic, less trustworthy, and raise negative emotions [41–43]. Such findings require further research to determine whether responses can truly be experienced as empathic when one knows their artificial origin; and whether such experiences differ in their relationship to treatment outcomes according to patients' theory of change.

## Conclusions

The advent of advanced AI technologies offers substantial benefits and potential enhancements to therapeutic practices, as well as greater accessibility for a wider population. Nevertheless, certain junctures within the therapeutic process may be particularly sensitive to the need for human rapport. We suggest that those points, which may be whole treatments or specific sessions, are times when empathy is especially needed. Although conversational AI can adeptly simulate empathic interactions, sometimes creating the impression of an empathy surpassing human capability [40], the essence of seeking empathy

transcends the mere reception of an ideal empathic response. It encompasses a longing for the genuine caring and emotional engagement from the individual offering support. The optimal path forward may lie in designing applications that facilitate therapist-AI partnerships, wherein AI systems could augment various facets of therapy—from initial intake and evaluation to, in certain instances, complete treatment modalities—while also consciously addressing the need for authentic human empathy, compassion, and care, when relevant for treatment success. However, most of our proposal is theoretical, and ultimately, we raise an empirical question, which should be evaluated in future studies. We also encourage industry professionals developing AI applications for mental health and those conducting research within this domain to remain mindful of these considerations.

## References

1. Weizenbaum J. ELIZA—a computer program for the study of natural language communication between man and machine. *Commun ACM*. 1966;9(1):36-45. doi:10.1145/365153.365168
2. Sullivan Y, Nyawa S, Fosso Wamba S. Combating Loneliness with Artificial Intelligence: An AI-Based Emotional Support Model. In: *Proceedings of the 56th Hawaii International Conference on System Sciences*. ; 2023. Accessed December 24, 2023. <https://hdl.handle.net/10125/103173>
3. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and Conversational Agents in Mental Health: A Review of the Psychiatric Landscape. *Can J Psychiatry Rev Can Psychiatr*. 2019;64(7):456-464. doi:10.1177/0706743719828977
4. Ly KH, Ly AM, Andersson G. A fully automated conversational agent for promoting mental well-being: A pilot RCT using mixed methods. *Internet Interv*. 2017;10:39-46. doi:10.1016/j.invent.2017.10.002
5. Inkster B, Sarda S, Subramanian V. An Empathy-Driven, Conversational Artificial Intelligence Agent (Wysa) for Digital Mental Well-Being: Real-World Data Evaluation Mixed-Methods Study. *JMIR MHealth UHealth*. 2018;6(11):e12106. doi:10.2196/12106

6. Inkster B, Kadaba M, Subramanian V. Understanding the impact of an AI-enabled conversational agent mobile app on users' mental health and wellbeing with a self-reported maternal event: a mixed method real-world data mHealth study. *Front Glob Womens Health*. 2023;4. Accessed August 20, 2023. <https://www.frontiersin.org/articles/10.3389/fgwh.2023.1084302>
7. Funk B, Sadeh-Sharvit S, Fitzsimmons-Craft EE, et al. A Framework for Applying Natural Language Processing in Digital Health Interventions. *J Med Internet Res*. 2020;22(2):e13855. doi:10.2196/13855
8. Ayers JW, Poliak A, Dredze M, et al. Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum. *JAMA Intern Med*. 2023;183(6):589. doi:10.1001/jamainternmed.2023.1838
9. Zaki J, Ochsner KN. The neuroscience of empathy: progress, pitfalls and promise. *Nat Neurosci*. 2012;15(5):675-680. doi:10.1038/nn.3085
10. Saxena A, Khanna A, Gupta D. Emotion Recognition and Detection Methods: A Comprehensive Survey. *J Artif Intell Syst*. 2020;2(1):53-79. doi:10.33969/AIS.2020.21005
11. USA, Shukla A. Utilizing AI and Machine Learning for Human Emotional Analysis through Speech-to-Text Engine Data Conversion. *J Artif Intell Cloud Comput*. Published online March 31, 2022:1-4. doi:10.47363/JAICC/2022(1)145
12. Chung LL, Kang J. "I'm Hurt Too": The Effect of a Chatbot's Reciprocal Self-Disclosures on Users' Painful Experiences. *Arch Des Res*. 2023;36(4):67-85. doi:10.15187/adr.2023.11.36.4.67
13. Yin Y, Jia N, Wakslak CJ. AI can help people feel heard, but an AI label diminishes this impact. *Proc Natl Acad Sci*. 2024;121(14):e2319112121. doi:10.1073/pnas.2319112121
14. Fuchs T. Empathy, Group Identity, and the Mechanisms of Exclusion: An Investigation into the Limits of Empathy. *Topoi*. 2019;38(1):239-250. doi:10.1007/s11245-017-9499-z
15. Cameron CD, Hutcherson CA, Ferguson AM, Scheffer JA, Hadjiandreou E, Inzlicht M. Empathy is hard work: People choose to avoid empathy because of its cognitive costs. *J Exp Psychol Gen*. 2019;148(6):962-976. doi:10.1037/xge0000595
16. Ferguson AM, Cameron CD, Inzlicht M. Motivational effects on empathic choices. *J Exp Soc Psychol*. 2020;90:104010. doi:10.1016/j.jesp.2020.104010
17. Perry A. AI will never convey the essence of human empathy. *Nat Hum Behav*. Published online July 20, 2023:1-2. doi:10.1038/s41562-023-01675-w

18. Elliott R, Bohart AC, Watson JC, Murphy D. Therapist empathy and client outcome: An updated meta-analysis. *Psychotherapy*. 2018;55(4):399-410. doi:10.1037/pst0000175
19. Rogers CR. *A Way of Being*. Houghton Mifflin; 1980.
20. Wampold BE, Imel ZE. *The Great Psychotherapy Debate: The Evidence for What Makes Psychotherapy Work*. Routledge; 2015.
21. DeRubeis RJ, Feeley M. Determinants of change in cognitive therapy for depression. *Cogn Ther Res*. 1990;14(5):469-482. doi:10.1007/BF01172968
22. Bordin ES. The generalizability of the psychoanalytic concept of the working alliance. *Psychother Theory Res Pract*. 1979;16(3):252-260. doi:10.1037/h0085885
23. Del Re AC, Flückiger C, Horvath AO, Wampold BE. Examining therapist effects in the alliance-outcome relationship: A multilevel meta-analysis. *J Consult Clin Psychol*. 2021;89(5):371-378. doi:10.1037/ccp0000637
24. Flückiger C, Del Re AC, Wampold BE, Horvath AO. The alliance in adult psychotherapy: A meta-analytic synthesis. *Psychotherapy*. 2018;55(4):316-340. doi:10.1037/pst0000172
25. Nienhuis JB, Owen J, Valentine JC, et al. Therapeutic alliance, empathy, and genuineness in individual adult psychotherapy: A meta-analytic review. *Psychother Res*. 2018;28(4):593-605. doi:10.1080/10503307.2016.1204023
26. Persons JB, Burns DD. Mechanisms of action of cognitive therapy: The relative contributions of technical and interpersonal interventions. *Cogn Ther Res*. 1985;9(5):539-551. doi:10.1007/BF01173007
27. Koelen JA, Vonk A, Klein A, et al. Man vs. machine: A meta-analysis on the added value of human support in text-based internet treatments ("e-therapy") for mental disorders. *Clin Psychol Rev*. 2022;96:102179. doi:10.1016/j.cpr.2022.102179
28. Musiat P, Johnson C, Atkinson M, Wilksch S, Wade T. Impact of guidance on intervention adherence in computerised interventions for mental health problems: a meta-analysis. *Psychol Med*. 2022;52(2):229-240. doi:10.1017/S0033291721004621
29. Werntz A, Amado S, Jasman M, Ervin A, Rhodes JE. Providing Human Support for the Use of Digital Mental Health Interventions: Systematic Meta-review. *J Med Internet Res*. 2023;25(1):e42864. doi:10.2196/42864
30. Zalaznik D, Strauss AY, Halaj A, et al. Patient alliance with the program predicts treatment outcomes whereas alliance with the therapist predicts adherence in internet-based therapy for panic disorder. *Psychother Res*. 2021;31(8):1022-1035. doi:10.1080/10503307.2021.1882712

31. Sedlakova J, Trachsel M. Conversational Artificial Intelligence in Psychotherapy: A New Therapeutic Tool or Agent? *Am J Bioeth.* 2023;23(5):4-13. doi:10.1080/15265161.2022.2048739
32. Asman O, Tal A, Barilan YM. Conversational Artificial Intelligence—Patient Alliance Turing Test and the Search for Authenticity. *Am J Bioeth.* 2023;23(5):62-64. doi:10.1080/15265161.2023.2191046
33. Middleton A, Woodward A, Gunn J, Bassilios B, Pirkis J. How do frequent users of crisis helplines differ from other users regarding their reasons for calling? Results from a survey with callers to Lifeline, Australia's national crisis helpline service. *Health Soc Care Community.* 2017;25(3):1041-1049. doi:10.1111/hsc.12404
34. Mishara BL, Côté LP, Dargis L. Systematic Review of Research and Interventions With Frequent Callers to Suicide Prevention Helplines and Crisis Centers. *Crisis.* 2023;44(2):154-167. doi:10.1027/0227-5910/a000838
35. Glikson E, Woolley AW. Human Trust in Artificial Intelligence: Review of Empirical Research. *Acad Manag Ann.* 2020;14(2):627-660. doi:10.5465/annals.2018.0057
36. Gould DJ, Dowsey MM, Glanville-Hearst M, et al. Patients' Views on AI for Risk Prediction in Shared Decision-Making for Knee Replacement Surgery: Qualitative Interview Study. *J Med Internet Res.* 2023;25(1):e43632. doi:10.2196/43632
37. Lucas GM, Rizzo A, Gratch J, et al. Reporting Mental Health Symptoms: Breaking Down Barriers to Care with Virtual Human Interviewers. *Front Robot AI.* 2017;4. Accessed September 7, 2023. <https://www.frontiersin.org/articles/10.3389/frobt.2017.00051>
38. Minerva F, Giubilini A. Is AI the Future of Mental Healthcare? *Topoi.* 2023;42(3):809-817. doi:10.1007/s11245-023-09932-3
39. Beatty C, Malik T, Meheli S, Sinha C. Evaluating the Therapeutic Alliance With a Free-Text CBT Conversational Agent (Wysa): A Mixed-Methods Study. *Front Digit Health.* 2022;4. Accessed January 10, 2024. <https://www.frontiersin.org/articles/10.3389/fdgth.2022.847991>
40. Lopes E, Jain G, Carlbring P, Pareek S. Talking Mental Health: a Battle of Wits Between Humans and AI. *J Technol Behav Sci.* Published online November 8, 2023. doi:10.1007/s41347-023-00359-6
41. Hohenstein J, Kizilcec RF, DiFranzo D, et al. Artificial intelligence in communication impacts language and social relationships. *Sci Rep.* 2023;13(1):5487. doi:10.1038/s41598-023-30938-9
42. Promberger M, Baron J. Do patients trust computers? *J Behav Decis Mak.* 2006;19(5):455-468. doi:10.1002/bdm.542

43. Rob Morris. We provided mental health support to about 4,000 people — using GPT-3. Here's what happened. X (formerly Twitter). Published January 6, 2023. Accessed September 14, 2023. <https://twitter.com/RobertRMorris/status/1611450197707464706>



## Supplementary Files

Untitled.

