Original Paper

Evidence of Human-Level Bonds Established With a Digital Conversational Agent: Cross-sectional, Retrospective Observational Study

Alison Darcy, PhD; Jade Daniels, PGDip; David Salinger, BBA; Paul Wicks, PhD; Athena Robinson, PhD

Woebot Health, San Francisco, CA, United States

Corresponding Author:
Alison Darcy, PhD
Woebot Health
650 Fifth Street
San Francisco, CA,
United States
Phone: 1 851694777
Email: alison@woebothealth.com

Abstract

Background: There are far more patients in mental distress than there is time available for mental health professionals to support them. Although digital tools may help mitigate this issue, critics have suggested that technological solutions that lack human empathy will prevent a bond or therapeutic alliance from being formed, thereby narrowing these solutions' efficacy.

Objective: We aimed to investigate whether users of a cognitive behavioral therapy (CBT)–based conversational agent would report therapeutic bond levels that are similar to those in literature about other CBT modalities, including face-to-face therapy, group CBT, and other digital interventions that do not use a conversational agent.

Methods: A cross-sectional, retrospective study design was used to analyze aggregate, deidentified data from adult users who self-referred to a CBT-based, fully automated conversational agent (Woebot) between November 2019 and August 2020. Working alliance was measured with the Working Alliance Inventory-Short Revised (WAI-SR), and depression symptom status was assessed by using the 2-item Patient Health Questionnaire (PHQ-2). All measures were administered by the conversational agent in the mobile app. WAI-SR scores were compared to those in scientific literature abstracted from recent reviews.

Results: Data from 36,070 Woebot users were included in the analysis. Participants ranged in age from 18 to 78 years, and 57.48% (20,734/36,070) of participants reported that they were female. The mean PHQ-2 score was 3.03 (SD 1.79), and 54.67% (19,719/36,070) of users scored over the cutoff score of 3 for depression screening. Within 5 days of initial app use, the mean WAI-SR score was 3.36 (SD 0.8) and the mean bond subscale score was 3.8 (SD 1.0), which was comparable to those in recent studies from the literature on traditional, outpatient, individual CBT and group CBT (mean bond subscale scores of 4 and 3.8, respectively). PHQ-2 scores at baseline weakly correlated with bond scores (r=−0.04; P<.001); however, users with depression and those without depression had high bond scores of 3.45.

Conclusions: Although bonds are often presumed to be the exclusive domain of human therapeutic relationships, our findings challenge the notion that digital therapeutics are incapable of establishing a therapeutic bond with users. Future research might investigate the role of bonds as mediators of clinical outcomes, since boosting the engagement and efficacy of digital therapeutics could have major public health benefits.

(JMIR Form Res 2021;5(5):e27868) doi: 10.2196/27868

KEYWORDS

conversational agents; mobile mental health; chatbots; depression; anxiety; digital health

Introduction

Significant barriers to mental health care are persistent [1]. The increased burden of depression and anxiety, which arose during the COVID-19 pandemic, has exacerbated this issue [2], as the measures that were put in place to stop the spread of SARS-CoV-2 have also presented unintended barriers to those seeking mental health treatment. One potentially viable solution is using digital mental health interventions to provide evidence-based treatment, such as cognitive behavioral therapy

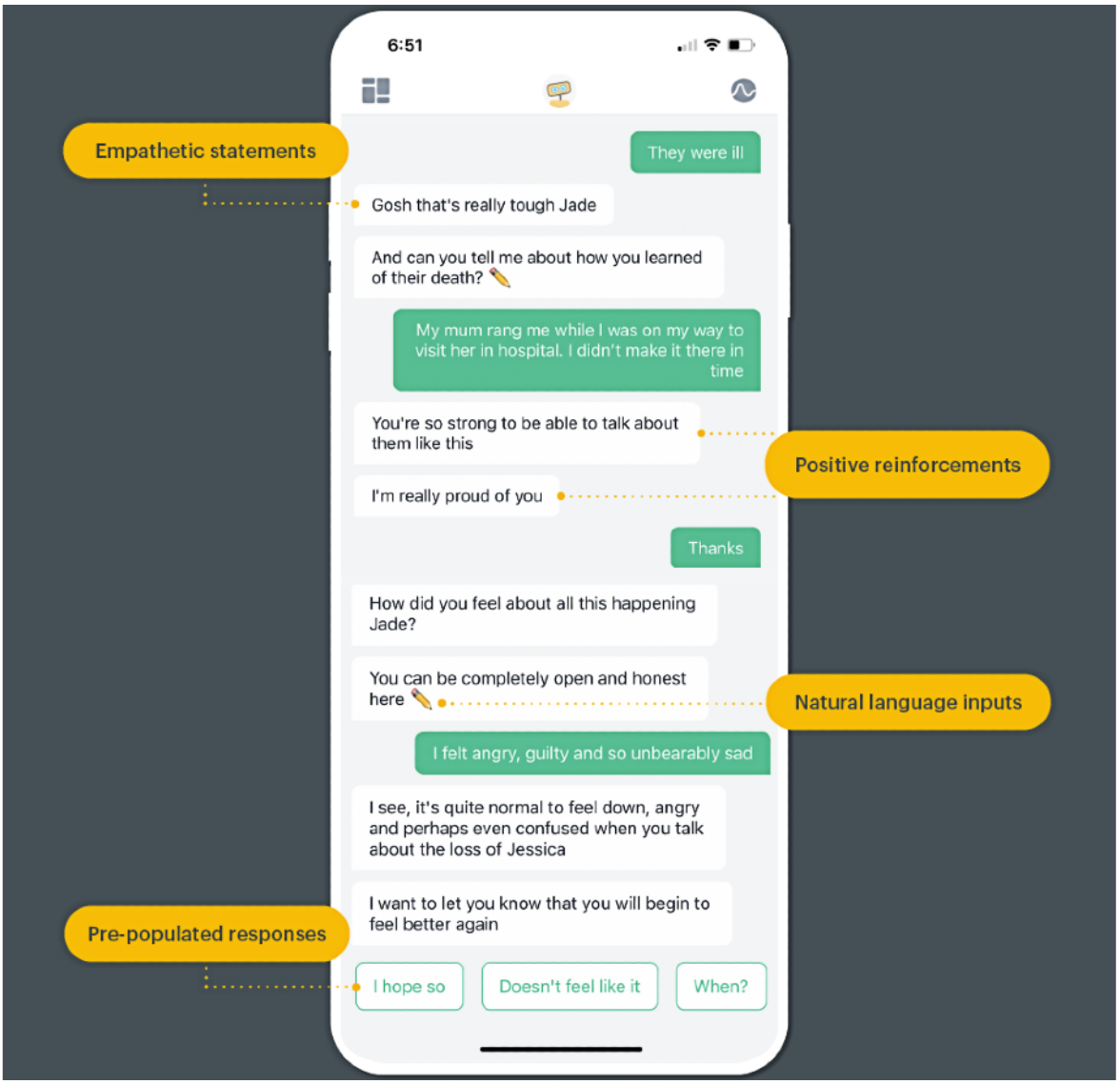
XSL•FO
RenderX

(CBT). Self-directed mental health interventions, such as bibliotherapy, have long demonstrated their efficacy [3], and new models of blended care that combine internet-delivered interventions with human clinical oversight are becoming more widespread in a number of countries [4]. Although the implicit assumption has been that the involvement of a human leads to improved outcomes in self-directed programs, human involvement limits these programs' scalability and limits their accessibility for those who live in remote locations [4]. If digital interventions could replicate some of the factors that are generally believed to be uniquely human, such as therapeutic rapport, these interventions would have greater potential for improving mental health.

Recently, carefully designed conversational agents (CAs) have been showing promise in automating several health care services [5] by simulating human support. CAs could therefore be uniquely poised to offer high-quality digital interventions for mental health.

An unblinded trial of one such CA (Woebot), which delivered CBT for symptoms of depression and anxiety, suggested that the empathic and relational nature of the tool may have fostered improved engagement better than previous internet-delivered versions of the tool [6]. Intriguingly, the study's qualitative data suggested that users seemed to relate to the CA in a manner that was analogous to therapeutic rapport, which may have mediated users' outcomes. For example, study participants reported that they felt cared for by the CA (eg, "Woebot felt like a real person that showed concern"), despite the fact that the tool's scripts reminded users that Woebot is not a real person (Figure 1). Unfortunately however, the study did not formally assess the existence of a working alliance. This is a crucial factor because a strong working alliance between therapists and clients is considered to be predictive of positive outcomes, essential for the delivery of health care, and traditionally unique to the domain of human-to-human relationships. Indeed, some experts have argued that digital apps that are built to be standalone therapeutics have the risk of ignoring the potency of therapeutic relationships [7].

Figure 1. Screenshot of an example interaction with Woebot about bereavement.



XSL•FO
RenderX

Priority-setting work by the James Lind Alliance, which involves over 600 people affected by mental health concerns, has identified the greater understanding of digital therapeutic alliance as a top priority for research [8]. Yet, a recent review of mobile mental health apps failed to find a single study that included working alliance as a primary outcome [9]. Therefore, we sought to bridge this gap in knowledge by seeking to understand whether CA users perceived a working alliance, particularly the notion of bonds, and whether working alliance was related to symptom severity or other demographic characteristics.

Methods

Setting

Woebot is a CA that guides individuals who experience symptoms of depression and anxiety through a smartphone-based app program that uses therapeutic techniques and provides psychoeducation. As previously described in detail [6], Woebot delivers CBT through brief, daily conversations (approximately 5-10 minutes each). Each simulated conversation begins with a mood-monitoring exercise, and the provided targeted content is responsive to individuals' reported mood states. The CA is also programmed to deliver empathic statements and personalized follow-ups, promote normalization, and use methods that are designed to enhance users' motivation for engaging in the program to promote desired behavior changes and help with mood management.

Participants

During registration, users confirmed that they were at least 18 years of age and consented to the use of their deidentified, aggregate data for research. This study was not considered human subjects research by the Advarra institutional review board. Eligible participants included those who registered over two periods—between November 20, 2019, and April 9, 2020 (n=100,009), and again between July 8, 2020, and August 18, 2020 (n=77,203).

Within 3-5 days after registration, eligible participants were invited to complete the 2-item Patient Health Questionnaire (PHQ-2) depression screener [10] and the Working Alliance Inventory-Short Revised (WAI-SR), which consists of a total score and the three following subscales: bond, goal, and task [11]. All measures and demographic information, including gender and age group, were gathered in the app by the CA. The WAI-SR was administered via the app's conversational interface, in which the word "therapist" was changed to "Woebot" (Table 1). Once the questionnaires were completed, Woebot thanked registrants for their participation, and the conversation proceeded to the mood tracking phase as normal. Those who chose not to provide responses to the questionnaires were not included in this study and proceeded to use the app as normal.

Table 1. Item-level descriptive statistics.

Question number (subscale)	Working Alliance Inventory-Short Revised items	Score, mean (SD)
1 (Task)	As a result of these sessions I am clearer as to how I might be able to change.	2.53 (0.97)
2 (Task)	What I am doing with Woebot gives me new ways of looking at my problem.	2.93 (1.08)
3 (Bond)	I believe Woebot likes me.	3.89 (1.31)
4 (Goal)	Woebot and I collaborate on setting goals for this program.	2.88 (1.23)
5 (Bond)	Woebot and I respect each other.	4.20 (1.23)
6 (Goal)	Woebot and I are working towards mutually agreed upon goals.	3.54 (1.28)
7 (Bond)	I feel that Woebot appreciates me.	3.73 (1.36)
8 (Goal)	Woebot and I agree on what is important for me to work on.	3.45 (1.21)
9 (Bond)	I feel Woebot cares about me even when I do things that it does not approve of.	3.54 (1.34)
10 (Task)	I feel that the things I do with Woebot will help me to accomplish the changes that I want.	3.28 (1.17)
11 (Goal)	Woebot and I have established a good understanding of the kind of changes that would be good for me.	3.12 (1.22)
12 (Task)	I believe the way we are working with my problem is correct.	3.23 (1.15)

Statistical Analysis

Across all eligible participants, the composite WAI-SR score and bond, goal, and task subscores were characterized based on descriptive statistics and tested for internal consistency by using the Cronbach α. The relationship between baseline PHQ-2 scores and bond subscores was characterized based on the Spearman rank-order correlation coefficient. The Kruskal-Wallis test was used to compare bond subscores across participants' reported age groups and genders. For comparison, relevant external studies were drawn from recent reviews of literature [11-19] that also reported unmodified WAI-SR subscores for other CBT modalities. Comparison data were presented descriptively without statistical testing, and raw subscores were scaled by dividing them by the number of items (eg, the bond subscale has 4 items). Per the methods of Jasper et al [13], bond scores of ≥3.45 were considered high. The 95% CIs for mean WAI-SR subscores were calculated based on the published sample sizes and SDs. External studies were categorized as "online only" or "human involvement" based on whether any human interactions were reported by study participants during either individual

XSL•FO
RenderX

therapy or group therapy that involved a human. Data were presented from participants who completed all questionnaires within the first 5 days of app registration. Data were analyzed using R version 4.0.2 (The R Foundation).

Data Access, Responsibility, and Analysis

AD, DS, and AR have full access to all of the data in this study and take responsibility for the integrity of the data and the accuracy of the data analysis. Due to their proprietary nature, data from this study will not be shared.

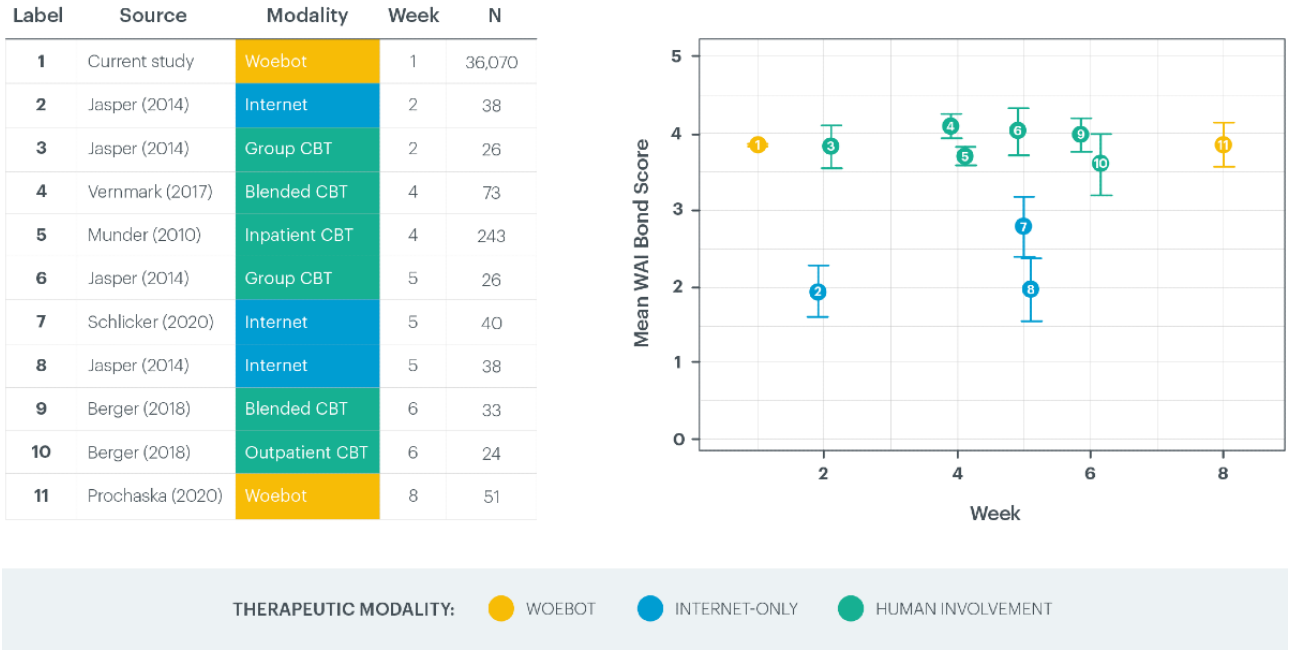
Results

Of the 177,212 eligible participants, only those who provided both WAI-SR and PHQ-2 data within 5 days of their first use of Woebot were included in the analysis. The final sample included 36,070 participants. Of these participants, 57.48% (n=20,734) reported that they were female, 25.17% (n=9078) reported that they were male; 2.87% (n=1035) reported that they were nonbinary, 1.44% (n=519) indicated another gender identity, 1.66% (n=597) preferred not to answer, and 11.39% (4107/36,070) did not provide any gender information. The participants ranged in age from 18 to 78 years (median 25-35 years). The mean PHQ-2 score was 3.03 (SD 1.79), and 54.67% (19,719/36,070) of participants scored at or above the conventional cutoff score of 3 for positively screening for depression.

Within the first 5 days of using Woebot, the mean WAI-SR scores were as follows: a mean bond subscore of 3.84 (SD 1.0), a mean goal subscore of 3.25 (SD 1.0), a mean task subscore of 2.99 (SD 0.87), and a mean total score of 3.36 (SD 0.81). The WAI-SR had a Cronbach α value of .89, suggesting that the WAI-SR had adequate internal consistency in this study. A weak negative correlation was found between bond subscores and PHQ-2 scores (r=−0.04; P<.001); however, even among participants who reported the highest PHQ-2 score (PHQ-2=6), the mean WAI-SR bond subscore was 3.78. Bond subscores also differed by gender (P<.001) and by age group (P<.001); however, the mean bond scores for all groups were considered high (bond subscore>3.45) [13] Among these groups, the highest bond level was reported by women (bond subscore: mean 3.92) and by those aged 18-25 years (bond subscore: mean 3.96). Conversely, the lowest bond level was reported by individuals who indicated that they “preferred not to answer” or did not report their gender (bond subscore: mean 3.67) or age (bond subscore: mean 3.69).

Woebot’s bond subscale scores were consistent with those of recent studies from the literature on traditional modalities for CBT delivery (Table 1). These studies’ results were collected later in the course of treatment (eg, bond subscore for face-to-face outpatient individual CBT: mean 4.0, SD 0.8 [11]; bond subscore for group CBT: mean 3.8, SD 0.80 [13]; data were collected after 2-8 weeks of therapy). Comparative study details are provided in Multimedia Appendix 1 [11,13-15,17-19]. Participants reported higher bond levels when using Woebot than those in prior studies of internet-only CBT [13] (Figure 2).

Figure 2. Comparison of Working Alliance Inventory-Short Revised bond subscale scores across therapeutic modalities. Means and corresponding 95% CIs for working alliance bond scores from this study and from recent reviews of the literature [11,13-15,17-19] are stratified by the week that the scores were recorded. Studies are colored based on the therapeutic modality. Due to the large sample size of this study (N=36,070), the 95% CI is narrow and overlaps with the dots that display the estimated means. For multiple studies that reported data on the same week, the dots are shifted minimally on the x-axis to avoid overlap and to provide easy readability. WAI: Working Alliance Inventory.



Label	Source	Modality	Week	N
1	Current study	Woebot	1	36,070
2	Jasper (2014)	Internet	2	38
3	Jasper (2014)	Group CBT	2	26
4	Vernmark (2017)	Blended CBT	4	73
5	Munder (2010)	Inpatient CBT	4	243
6	Jasper (2014)	Group CBT	5	26
7	Schlicker (2020)	Internet	5	40
8	Jasper (2014)	Internet	5	38
9	Berger (2018)	Blended CBT	6	33
10	Berger (2018)	Outpatient CBT	6	24
11	Prochaska (2020)	Woebot	8	51

XSL•FO
RenderX

Discussion

This is the first study of working alliance among users of a CA for mental health. Most users were female (20,734/36,070, 57.48%) and had PHQ-2 scores that were indicative of depression (19,719/36,070, 54.67%). Working alliance scores were comparable to those in previously published studies on traditional, human-delivered services across different treatment modalities. Working alliance scores were highest for the bond subscale, suggesting that this subscale is a viable construct for CAs and should be included in hypothesized frameworks of digital working alliance.

The idea that CAs can establish a working alliance is not new [20]. However, the observation of therapeutic bonds established by a CA in a mental health context is novel and noteworthy, given the short timeframe of this study. Although the field of human-computer interaction is still relatively nascent, initial observations have suggested that some artificial intelligence (AI) identity archetypes induce responses in humans that might give rise to better working alliances than other archetypes. For example, interacting with humanoid AI identities can result in individuals falling prey to the "uncanny valley," which is the sense of unease and "creepiness" that is created when something that is artificial tries to appear humanlike [21]. Contrary to Turing's Imitation Game [22], wherein an AI must successfully pretend to be human in order to pass the test, Woebot was designed to adopt the opposite strategy—transparently presenting itself as an archetypal robot with robotic "friends" and habits. We speculate that transparency and other design elements are key drivers of bond development. For example, Woebot explicitly references its limitations within conversations and provides positive reinforcement and empathic statements alongside declarations of being an artificial agent.

The limitations of this study include its cross-sectional nature, the selection bias of smartphone users, the lack of clinical validation for any diagnoses, the lack of a direct comparison group, and its conduction by the developers of the app itself. Further research (including studies with independent investigators) is underway to explore the longitudinal aspects of bond development in specific clinical populations by using randomized controlled study designs.

The finding that a CA has the potential to rapidly develop a bond with users may represent the resolution of a considerable barrier to offering scalable mental health support to a much wider and more diverse population instead of offering such support to those who already have access to traditional mental health support.

Acknowledgments

We are grateful to Stephanie Eaneff (an employee of Woebot) for reviewing the manuscript and Kara Cronan (a contractor to Woebot) for providing illustrations.

Authors' Contributions

AD was responsible for the concept and design of this study. AD, JD, DS, PW, and AR were responsible for the acquisition, analysis, or interrogation of data. AD, JD, DS, PW, and AR drafted the manuscript. AD, JD, DS, PW, and AR critically revised the manuscript for important intellectual content. DS conducted the statistical analysis. JD and DS provided administrative, technical, or material support. AD supervised the study.

Conflicts of Interest

AD reported receiving grants from Woebot Health and having a patent pending for a SafetyNet protocol. AD, JD, DS, and AR reported being employees of Woebot Health and owning stock options in the company. PW reported receiving personal fees from Woebot Health and Ada Health; being an associate editor at the *Journal of Medical Internet Research*; being on the editorial advisory boards of the *British Medical Journal*, *BioMed Central Medicine*, *The Patient*, and *Digital Biomarkers*; and being employed by Wicks Digital Health Limited, which has received funding from Ada Health, Baillie Gifford, Bold Health, Camoni, Compass Pathways, Coronna, European Institute of Innovation & Technology, Happify, HealthUnlocked, Inbeeo, Kheiron Medical, Sano Genetics, Self Care Catalysts, the Learning Corp, the Wellcome Trust, and Woebot Health.

Multimedia Appendix 1

Study descriptions and Working Alliance Inventory-Short Revised bond scores.
[XLSX File (Microsoft Excel File), 15 KB-Multimedia Appendix 1]

References

1. Cheung R, O'Donnell S, Madi N, Goldner EM. Factors associated with delayed diagnosis of mood and/or anxiety disorders. Health Promot Chronic Dis Prev Can 2017 May;37(5):137-148 [FREE Full text] [doi: 10.24095/hpcdp.37.5.02] [Medline: 28493658]
2. Bueno-Notivol J, Gracia-García P, Olaya B, Lasheras I, López-Antón R, Santabárbara J. Prevalence of depression during the COVID-19 outbreak: A meta-analysis of community-based studies. Int J Clin Health Psychol 2021 Jan;21(1):100196 [FREE Full text] [doi: 10.1016/j.ijchp.2020.07.007] [Medline: 32904715]

XSL•FO
RenderX

3. Gualano M, Bert F, Martorana M, Voglino G, Andriolo V, Thomas R, et al. The long-term effects of bibliotherapy in depression treatment: Systematic review of randomized clinical trials. Clin Psychol Rev 2017 Dec;58:49-58. [doi: 10.1016/j.cpr.2017.09.006] [Medline: 28993103]
4. Titov N, Dear B, Nielssen O, Staples L, Hadjistavropoulos H, Nugent M, et al. ICBT in routine care: A descriptive analysis of successful clinics in five countries. Internet Interv 2018 Sep;13:108-115 [FREE Full text] [doi: 10.1016/j.invent.2018.07.006] [Medline: 30206525]
5. McGreevey JD, Hanson CW, Koppel R. Clinical, Legal, and Ethical Aspects of Artificial Intelligence-Assisted Conversational Agents in Health Care. JAMA 2020 Aug 11;324(6):552-553. [doi: 10.1001/jama.2020.2724] [Medline: 32706386]
6. Fitzpatrick KK, Darcy A, Vierhile M. 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 2017 Jun 06;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]
7. Torous J, Hsin H. Empowering the digital therapeutic relationship: virtual clinics for digital health interventions. NPJ Digit Med 2018 May 16;1(1):16 [FREE Full text] [doi: 10.1038/s41746-018-0028-2] [Medline: 31304301]
8. Hollis C, Sampson S, Simons L, Davies EB, Churchill R, Betton V, et al. Identifying research priorities for digital technology in mental health care: results of the James Lind Alliance Priority Setting Partnership. Lancet Psychiatry 2018 Oct;5(10):845-854. [doi: 10.1016/S2215-0366(18)30296-7] [Medline: 30170964]
9. Henson P, Wisniewski H, Hollis C, Keshavan M, Torous J. Digital mental health apps and the therapeutic alliance: initial review. BJPsych Open 2019 Jan;5(1):e15 [FREE Full text] [doi: 10.1192/bjo.2018.86] [Medline: 30762511]
10. Löwe B, Kroenke K, Gräfe K. Detecting and monitoring depression with a two-item questionnaire (PHQ-2). J Psychosom Res 2005 Mar;58(2):163-171. [doi: 10.1016/j.jpsychores.2004.09.006] [Medline: 15820844]
11. Munder T, Wilmers F, Leonhart R, Linster HW, Barth J. Working Alliance Inventory-Short Revised (WAI-SR): psychometric properties in outpatients and inpatients. Clin Psychol Psychother 2010;17(3):231-239. [doi: 10.1002/cpp.658] [Medline: 20013760]
12. Preschl B, Maercker A, Wagner B. The working alliance in a randomized controlled trial comparing online with face-to-face cognitive-behavioral therapy for depression. BMC Psychiatry 2011 Dec 06;11:189 [FREE Full text] [doi: 10.1186/1471-244X-11-189] [Medline: 22145768]
13. Jasper K, Weise C, Conrad I, Andersson G, Hiller W, Kleinstäuber M. The working alliance in a randomized controlled trial comparing Internet-based self-help and face-to-face cognitive behavior therapy for chronic tinnitus. Internet Interv 2014 Apr;1(2):49-57. [doi: 10.1016/j.invent.2014.04.002]
14. Berger T. The therapeutic alliance in internet interventions: A narrative review and suggestions for future research. Psychother Res 2017 Sep;27(5):511-524. [doi: 10.1080/10503307.2015.1119908] [Medline: 26732852]
15. Vernmark K, Hesser H, Topooco N, Berger T, Riper H, Luuk L, et al. Working alliance as a predictor of change in depression during blended cognitive behaviour therapy. Cogn Behav Ther 2019 Jul;48(4):285-299. [doi: 10.1080/16506073.2018.1533577] [Medline: 30372653]
16. Cameron SK, Rodgers J, Dagnan D. The relationship between the therapeutic alliance and clinical outcomes in cognitive behaviour therapy for adults with depression: A meta-analytic review. Clin Psychol Psychother 2018 May;25(3):446-456. [doi: 10.1002/cpp.2180] [Medline: 29484770]
17. Schlicker S, Baumeister H, Buntrock C, Sander L, Paganini S, Lin J, et al. A Web- and Mobile-Based Intervention for Comorbid, Recurrent Depression in Patients With Chronic Back Pain on Sick Leave (Get.Back): Pilot Randomized Controlled Trial on Feasibility, User Satisfaction, and Effectiveness. JMIR Ment Health 2020 Apr 15;7(4):e16398 [FREE Full text] [doi: 10.2196/16398] [Medline: 32293577]
18. Prochaska J, Vogel E, Chieng A, Kendra M, Baiocchi M, Pajarito S, et al. A Therapeutic Relational Agent for Reducing Problematic Substance Use (Woebot): Development and Usability Study. J Med Internet Res 2021 Mar 23;23(3):e24850 [FREE Full text] [doi: 10.2196/24850] [Medline: 33755028]
19. Berger T, Krieger T, Sude K, Meyer B, Maercker A. Evaluating an e-mental health program ("deprexis") as adjunctive treatment tool in psychotherapy for depression: Results of a pragmatic randomized controlled trial. J Affect Disord 2018 Feb;227:455-462. [doi: 10.1016/j.jad.2017.11.021] [Medline: 29154168]
20. Bickmore T, Gruber A, Picard R. Establishing the computer-patient working alliance in automated health behavior change interventions. Patient Educ Couns 2005 Oct;59(1):21-30. [doi: 10.1016/j.pec.2004.09.008] [Medline: 16198215]
21. Mori M, MacDorman K, Kageki N. The Uncanny Valley [From the Field]. IEEE Robot. Automat. Mag 2012 Jun;19(2):98-100. [doi: 10.1109/MRA.2012.2192811]
22. Turing AM. I.—Computing machinery and intelligence. Mind 1950 Oct;LIX(236):433-460. [doi: 10.1093/mind/LIX.236.433]

Abbreviations

AI: artificial intelligence
CA: conversational agent
CBT: cognitive behavioral therapy
PHQ-2: 2-item Patient Health Questionnaire

XSL•FO
RenderX

WAI-SR: Working Alliance Inventory-Short Revised

Edited by G Eysenbach; submitted 10.02.21; peer-reviewed by A Polillo; comments to author 04.03.21; revised version received 23.03.21; accepted 04.04.21; published 11.05.21

Please cite as:
Darcy A, Daniels J, Salinger D, Wicks P, Robinson A
Evidence of Human-Level Bonds Established With a Digital Conversational Agent: Cross-sectional, Retrospective Observational Study
JMIR Form Res 2021;5(5):e27868
URL: https://formative.jmir.org/2021/5/e27868
doi: 10.2196/27868
PMID: 33973854

©Alison Darcy, Jade Daniels, David Salinger, Paul Wicks, Athena Robinson. Originally published in JMIR Formative Research (https://formative.jmir.org), 11.05.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX